Using Chemical Structure and Inocula Characteristics to Predictively Model Biodegradation Rate

Kishor Acharya, B. Pharm, MSc

A thesis submitted to Newcastle University for partial fulfilment of the requirements of the degree of Doctor of Philosophy within the School of Engineering

School of Engineering
Newcastle University
Newcastle upon Tyne
NE1 7RU

September 2018

Abstract

Predictive biodegradation models [i.e. Quantitative Structure Biodegradation Relationship (QSBR) models] might be used as an alternative to current regulatory biodegradation tests to predict chemical persistence. Current models are mostly based on the results derived from regulatory Ready Biodegradability Tests (RBTs), which are highly variable and were not designed to provide half-life data and therefore fundamentally undermines efforts to reliably predict chemical persistence. Improvement to existing approaches for developing and verifying predictive models and their reliability, respectively, have been proposed, and the use of functional gene and 16S rRNA amplicon sequencing techniques towards identifying and quantifying the putative chemical degraders have been studied. Several QSBR models for aromatic chemicals were developed according to OECD principles. Models for mono-aromatic chemicals were verified and calibrated with experimentally determined rates (both from pure culture and natural mixed communities). Traditional test methods were combined with functional genes and 16S amplicon sequence analyses to develop a relationship between rate, chemical concentration and competent putative chemical degrader abundance. QSBR models for mono-aromatic chemicals were stable ($R^2 = 0.8924$), robust ($Q^2_{LOO} = 0.8718$) and had good predictive ability ($Q^2_{F1} = 0.8829$, $Q^2_{F2} = 0.8835$, and $Q^2_{F3} = 0.9178$). In these models, biodegradation rates were associated with electronic, lipophilic and steric descriptors, and thus provided information on the mechanisms of different rate-limiting steps associated with the biodegradation process. However, all the variation in biodegradation rates cannot be explained by the structure alone, the prevailing environmental conditions have a significant role in determining the extent of chemical degradation. Biodegradation rates (k) of chemicals in natural mixed communities were significantly correlated with the ratio of abundance of initial putative degrader abundances (X_0) and the starting chemical concentration (C_0) (Pearson correlation coefficient (r) > 0.9 and p-value < 0.05). Predictive models developed by relating k with X_0 and C_0 reliably predicted the rate of studied chemicals. Experimentally determined rates further formed the basis towards calibrating the developed QSBR models. The molecular analysis revealed that majority of identified putative chemical degraders were rare taxa, and their enrichment did not necessarily influence the overall biomass count of the microbial community, and therefore biodegradation models that only consider the overall biomass would not account for the kind of relationships found in this study. Application of 16S amplicon sequencing and functional gene analyses techniques in biodegradation studies will help in depth screening of diversity and function of microbial community in an inoculum and enables better understanding of biodegradation outcomes.

Declaration

I hereby certify that this work is my own, except otherwise acknowledged, and that is has not been submitted for fulfilment of a degree at this or any other university.

Kishor Acharya

Contents

Abstract ..... **i**

Declaration ..... **ii**

List of Tables ..... **vii**

List of Figures ..... **x**

Acknowledgements ..... **xiv**

Chapter 1. General Introduction **2**

Chapter 2. Literature Review **6**

2.1 Human health and environmental implication of organic chemicals **6**

2.2 Fate of organic chemicals in the environment **7**

2.3 Chemical persistence and its assessment **8**

2.3.1 Biodegradation Kinetics ***10***

2.4 Biodegradation, biodegradation tests and risk assessment **12**

2.4.1 Limitations of OECD tests ***16***

2.5 Predictive modelling, an alternative to biodegradation test **17**

2.5.1 Quantitative Structure Activity Relationships (QSARs) ***18***

2.5.2 Different Approaches for Modelling Biodegradation ***20***

2.5.3 Conditions for using QSAR results ***23***

2.6 Biodegradation processes - a 'black box' for ecological modelling programs **25**

2.6.1 Microbial ecology in biodegradation testing ***25***

2.6.2 Role of catabolic genes in aerobic biodegradation of chemicals ***26***

2.7 Research Gaps **27**

2.8 Aims and Objectives **28**

Chapter 3. A statistical modelling approach to predict biodegradation rates of aromatic chemicals. **30**

3.1 Abstract **30**

3.2 Introduction **30**

3.3 Methods **32**

3.3.1 Chemical Selection and Molecular Descriptors Choice ***32***

3.3.2 Endpoint for QSBR model ***32***

3.3.3 Screening of Chemicals for QSBR model development ***33***

3.3.4 Dataset Splitting ***33***

3.3.5 QSBR Model Development and Validation ***34***

3.4 Result and Discussion **37**

3.4.1 QSBR models ***37***

3.4.2 QSBR model for simple aromatic chemicals (Group 1) *39*
3.4.3 QSBR model descriptors and their mechanistic interpretation *41*
3.4.4 Mechanism of Aerobic Biodegradation *42*
3.4.5 Implications for predicting biodegradation half-lives for fate and hazard assessment ... *45*
3.4.6 Future outlook *46*

Chapter 4. A Biodegradation Test Method for Estimating Biodegradation Rates towards Evaluating QSBR Models **48**

4.1 Abstract **48**
4.2 Introduction **48**
4.3 Experimental Procedures **50**
4.3.1 Chemicals and Bacterial strains *50*
4.3.2 Experimental System *51*
4.3.3 Biodegradation assays *52*
4.3.4 Carbon mass balance *53*
4.3.5 Modelling biodegradation and estimation of kinetic parameters *54*
4.4 Results and Discussion **55**
4.4.1 Batch biodegradation of chemicals *55*
4.4.2 Biodegradation kinetics and parameters estimation *55*
4.4.3 Carbon mass balance and stoichiometry to assess uncertainty in the biodegradation experiments *59*
4.4.4 Verification and calibration of a QSBR Model *64*
4.5 Conclusions **66**

Chapter 5. Development and application of molecular tools to study the biodegradation of mono-aromatic chemicals in activated sludge **68**

5.1 Abstract **68**
5.2 Introduction **68**
5.3 Material and Methods **70**
5.3.1 Selection of catabolic genes and reference microorganisms *70*
5.3.2 Phylogenetic Analysis *70*
5.3.3 Primer design *71*
5.3.4 Bacterial strain and culture conditions *71*
5.3.5 DNA extraction *71*
5.3.6 RNA Extraction *75*
5.3.7 PCR Analysis *75*
5.3.8 Quantitative real time PCR analysis for quantification of catabolic genes *76*
5.3.9 Catabolic gene standards for calibration of real-time PCR *77*
5.3.10 Quantification of gene expression *77*

5.3.11 Biodegradation assays for phenol and 2,4-DCP ... *79*
5.3.12 Wastewater treatment plant sampling ... *79*
5.3.13 Biodegradation Assays ... *80*
5.3.14 Quantification of total bacteria in activated sludge using flow cytometery ... *81*
5.4 Results ... **81**
5.4.1 Specificity of the primer-sets ... *81*
5.4.2 Quantitative real-time PCR (qPCR) optimization ... *82*
5.4.3 Biodegradation assay ... *84*
5.5 Discussion ... **91**
5.5.1 Primer design ... *92*
5.5.2 qPCR Assay ... *93*
5.5.3 Trial biodegradation assays ... *93*
5.5.4 Gene abundance and expression ... *94*
5.6 Conclusion ... **96**

Chapter 6. Can initial competent degrader abundance and chemical concentration be used to predict biodegradation rate of aromatic chemicals in activated sludge? ... **98**
6.1 Abstract ... **98**
6.2 Introduction ... **98**
6.3 Materials and Methods ... **101**
6.3.1 Waste water treatment plant sampling ... *101*
6.3.2 Dilution of activated sludge ... *101*
6.3.3 Dilution of activated sludge with UV treated activated sludge ... *102*
6.3.4 Experimental design for Biodegradation assay ... *104*
6.3.5 Biodegradation assays ... *105*
6.3.6 DNA extraction and gene quantification ... *106*
6.4 Results and Discussion ... **106**
6.4.1 DOE results for phenol and 2,4-DCP biodegradation assays ... *106*
6.4.2 Multi-regression model ... *112*
6.4.3 Effect of initial chemical concentration and specific degrader numbers on rate ... *117*
6.5 Conclusion ... **120**

Chapter 7. Use of 16S rRNA amplicon sequencing and functional gene analyses to reveal the associations between putative degraders and chemical degradation in activated sludge. ... **122**
7.1 Abstract ... **122**
7.2 Introduction ... **122**
7.3 Material and Methods ... **125**
7.3.1 Chemicals, activated sludge sampling and dilution ... *125*

7.3.2 Batch biodegradation experiments *125*
7.3.3 DNA extraction, total catabolic gene quantification and PCR amplification (library preparation) *126*
7.3.4 PCR amplicon library preparation *126*
7.3.5 Microbial community structure analysis *127*
7.3.6 Statistical Analysis *127*
7.4 Result and Discussion **128**
7.4.1 Summary of biodegradation assay results for test chemicals *128*
7.4.2 Variation in functional gene and total bacterial abundances during test chemical biodegradation *129*
7.4.3 Microbial community structure *136*
7.4.4 OTU functions in test chemical biodegradation *137*
7.4.5 Distribution of rare and abundant taxa in microbial community *142*
7.5 Implication of 16S rRNA amplicon sequencing and functional gene analyses: rare taxa identification and screening of inocula **146**
7.6 Implication of quantifying key catabolic genes in predicting the biodegradation kinetics **153**
7.7 Conclusion **154**
7.8 Verification and calibration of QSBR model **155**
Chapter 8. Conclusions, broader implication and recommendations for future work 159
8.1 Conclusions **159**
8.2 Broader implication of the current research **160**
8.3 Future outlook **161**
Reference **163**
Appendices **183**

List of Tables

Table 1-1 Overview of the environmental exposures associated with chronic non-communicable diseases.....4
Table 1-2 Different chemicals related problems in the environment 4
Table 2-1 The persistency criteria set by different regulatory bodies [Adapted from ; (Pavan and Worth, 2008)].....9
Table 2-2 Required test data of interest for the integrated testing strategy (ITS) on degradation 17
Table 2-3 List of the most commonly used databases for biodegradation studies 20
Table 2-4 Six different Biowin models included in Biodegradation Probability Program (BIODEG) [Adapted from; (Pavan and Worth, 2008)] 22
Table 2-5 Examples of different QSAR models. The nature of QSBR model (Qualitative or Quantitative) are indicated, different kinds of independent variables used by the models are mentioned with techniques used for modelling.....24
Table 3-1 Summary of the Best QSBR model for each of the selected groups of chemicals 38
Table 3-2 Summary of the 10 best models with 4 descriptors......40
Table 3-3 Model statistics of 3 descriptors QSBR model......40
Table 4-1 List of chemicals, their known degrader strains and their designation 51
Table 4-2 Estimated Monod model parameters values for aerobic biodegradation of the test chemicals by known degraders.....59
Table 4-3 Comparison of theoretically estimated biomass and CO_2 amount for each test chemical biodegradation with the experimentally observed amounts......62
Table 4-4 Summary table for correlation (Pearson correlation) values......63
Table 5-1 Key catabolic genes involved in phenol and 2,4-DCP aerobic biodegradation pathways and bacterial referene strains used in this study.....70
Table 5-2 List of primers used for the amplification of gene fragments coding for PH, C23D, tfdC and tfdB using PCR and real-time PCR.....74
Table 5-3 Summary of an optimum annealing temperature and efficiency of the primers used in this study for qPCR and RT-qPCR assays 83
Table 5-4 List of WWTPs, description, properties, abundance of initial phenol catabolic genes and first-order phenol biodegradation rates in trial experiments.....85
Table 5-5 List of WWTPs, description, properties, abundance of initial 2,4-DCP catabolic genes and first-order 2,4-DCP biodegradation rates in trial experiments.....85

Table 5-6 List of WWTPs, description, properties, abundance of initial phenol and 2,4-DCP catabolic genes and first order biodegradation rate in final experiments..... 85
Table 5-7 Ratio of total catabolic gene transcripts and total catabolic genes over time during the biodegradation of phenol and 2,4-DCP in activated sludge sampled from WWTP1 and WWTP4, respectively..... 90
Table 6-1 Biodegradation model for mineralization kinetics derived from integrated Monod Equations with the variable of substrate concentration and inocula concentration..... 99
Table 6-2 Factorial design experiments in biodegradation studies..... 101
Table 6-3 First order biodegradation rates of 2,4-DCP in UV-treated and untreated activated sludge...... 103
Table 6-4 Level of two variables in coded (+1/-1) and un-coded form for the degradation of phenol and 2,4-DCP using activated sludge (AS) grab samples as inocula....... 104
Table 6-5 Experimental conditions (for both 2,4-DCP and phenol assays) created with a factorial design (2^2) for two variables...... 104
Table 6-6 Experimental conditions of reactors for phenol DOE validation experiments. The dilution of inocula were performed with UV-sterilized activated sludge (AS)...... 105
Table 6-7 Experimental condition of reactors for 2,4-DCP DOE validation experiments. The dilutions of activated sludge (AS) inocula were performed with OECD mineral medium...... 105
Table 6-8 Summary of the initial catabolic gene copy number per mL of activated sludge (AS) in different reactors used for phenol DOE and validation experiments, observed and model predicted phenol biodegradation rates for each reactor (Top; PH gene and Bottom; C23D gene)...... 110
Table 6-9 Summary of the initial catabolic gene copy numbers per mL of activated sludge (AS) in different reactors used for 2,4-DCP DOE experiments and validation experiments, observed and model predicted 2,4-DCP biodegradation rates for each reactor (Top; *tfdC* gene and bottom; *tfdB* gene)....... 111
Table 6-10 Analysis of variance for the multiple regression linear model (Equation 1A) for phenol from the factorial design experiments based of PH gene copy numbers. 115
Table 6-11 Analysis of variance for multiple regression linear model (equation 2) for 2,4-DCP factorial design experiments based of tfdB gene copies number..... 116
Table 6-12 Correlation analysis between the first order biodegradation rate of test chemicals with different parameters (X_0, C_0 and X_0/C_0)..... 118

Table 7-1 Experimental condition for the biodegradation assays of four test chemicals (phenol, m-cresol, 4-chlorophenol and 2,4-DCP) 125
Table 7-2. Experimentally estimated biodegradation rates of four test chemicals for the selected experimental conditions.......128
Table 7-3 Summary statistics of univariate regression analysis between first order biodegradation rates of test chemicals and the ratio of X_0 (initial putative competent degrader numbers) and C_0 (initial spiked chemical concentration). 2,4-DCP and 4-CP represent 2,4-dichlorophenol and 4-chlorophenol respectively. 129
Table 7-4 Summary table for Monod no – growth model fits for each chemical. R^2 value reported is relationship between the observed chemical concentration and the model predicted chemical concentration. 134
Table 7-5 Summary table for Monod no – growth model fits for each chemical. R^2 value reported is relationship between the observed chemical concentration and the model predicted chemical concentration.. 134
Table 7-6 Summary statistics of univariate regression analysis between final and initial relative abundance of observed genera in the degradation assays performed for four different test chemicals (5 mg/L) at two different inocula concentration. 146
Table 7-7 Summary of statistically significant treatment effects [time zero undiluted activated sludge inocula (i.e.before spiking the test chemicals) versus final inocula (i.e. end of biodegradation assays), G-test, $p<0.05$] for OTU identified at genus level, where member of genus reportedly degrade test chemicals and aromatic hydrocarbons......149
Table 7-8 Summary of statistically significant treatment effects [time zero diluted activated sludge inocula (i.e.before spiking the test chemicals) versus final inocula (i.e. end of biodegradation assays), G-test, $p<0.05$] for OTU identified at genus level, where member of genus reportedly degrade test chemicals and aromatic hydrocarbons. 151
Table 7-9 Summary statistics of univariate regression analysis between experimentally determined biodegradation rates at different studied conditions including pure culture experiments and the biodegradation rates used in previously developed QSBR models for four test chemicals.. 156

List of Figures

Figure 1-1 Chemical industry output for (A) Developed Regions and (B) Developing Regions & Countries with economies in transition s as categorized by UN Statistics Division. 3
Figure 2-1 Transport and fate of chemicals entering the aquatic environment 8
Figure 2-2 Effect of substrate concentration on kinetics of reaction [Reproduced from : (Battersby, 1990)] 12
Figure 2-3 The first step in the persistence assessment of chemical under REACH [Reproduced from; (Martin, 2014)] P/vP = persistent/very Persistent 13
Figure 2-4 Schematic overview of using degradation data in classification and Labelling 14
Figure 2-5 Summary of decision scheme on degradation for three different regulatory needs 15
Figure 3-1 Detailed procedure of developing QSBR (adopted from (Mikolajczyk et al., 2015)) 36
Figure 3-2 (A) Plot of BIOWIN3 derived rates versus model calculated values for the full model based on three descriptors.. 41
Figure 3-3 Williams plot for the applicability domain of the full model based on three descriptors 41
Figure 4-1 Plot of BIOWIN3 derived rates versus model calculated rates. 50
Figure 4-2 Illustration of the glass bottle reactor used to conduct biodegradation studies of test chemicals 52
Figure 4-3 Chemical concentrations, biomass growth and headspace CO_2 concentration profiles obtained during biodegradation experiments with known degraders at room temperature for phenol, m-cresol, 4-chlorophenol and 2,4-dichlorophenol and toluene. 57
Figure 4-4 Clustered stacked column chart summarizing the mass balance result for each test chemical biodegradation in reactors.. 62
Figure 4-5 Relationship between the experimentally determined biodegradation rates and the biodegradation rates used in previously developed QSBR models for five test chemicals. 65
Figure 5-1 Unrooted phylogenetic neighbour-joining tree for; [A] 2,4-dichlorophenol hydroxylase, [B] Phenol hydroxylase, and [C] catechol-2,3-dioxygenase genes based on nucleotide sequences taken from the GeneBank. 73
Figure 5-2 Biodegradation of phenol in activated sludge sampled from different WWTPs. 84

Figure 5-3 Biodegradation of phenol in activated sludge. 86
Figure 5-4 Biodegradation of 2,4-DCP in activated sludge. 87
Figure 5-5 Total bacterial counts and phenol catabolic gene (C23D and PH) copies over time for the phenol biodegradation experiment conducted with activated sludge inocula sampled from WWTP1.. 88
Figure 5-6 Total bacterial counts and 2,4-DCP catabolic gene (tfdB and tfdC) copies over time for the 2,4-DCP biodegradation experiments conducted with activated sludge inocula sampled from WWTP4......89
Figure 5-7 Phenol removal and expression profile of PH and C23D genes over time during the biodegradation of phenol using an activated sludge inoculum sampled from WWTP1......91
Figure 5-8 2,4-DCP removal and expression profile of tfdB and tfdC genes over time during the biodegradation of 2,4-DCP using an inoculum of activated sludge sampled from WWTP4......91
Figure 5-9 Univariate regression analysis between first order biodegradation rate of phenol and initially observed PH (bottom) and C23D (top) catabolic genes for phenol biodegradation in activated sludge inocula from different WWTPs......94
Figure 6-1Viable cell counts (on R2A agar plate) at different time intervals during UV sterilization of activated sludge grab sample (A), and phenol biodegradation using UV - treated and untreated activated sludge grab samples as inocula (B)......103
Figure 6-2. 2,4-DCP biodegradation in autoclaved (Control), UV-treated and untreated activated sludge......103
Figure 6-3 Phenol biodegradation profile for experimental conditions created with a factorial design (2^2) for the training experiments......107
Figure 6-4 Phenol biodegradation profiles for DOE validation experiments. The figure chart title shows the experimental condition of each reactor......108
Figure 6-5 Phenol biodegradation profile in UV treated activated sludges from different WWTPs, which acted as a control reactor......109
Figure 6-6 Relationship between the experimentally determined biodegradation rates and the model predicted biodegradation rates for (A) phenol with PH gene and (B) 2,4-DCP with tfdB gene, (C) phenol with C23D gene and (D) 2,4-DCP with tfdC gene. 114
Figure 6-7 Main effect (A and C) and interaction plots (B and D) for first order biodegradation rates in relation to intial phenol (with PH gene) and 2,4-DCP (tfdB gene) degrader concentrations. 118

Figure 6-8 Relationship between first order biodegradation rates (k) of test chemicals (phenol and 2,4-DCP) with different parameters (X_0, C_0 and X_0/C_0).. 119
Figure 7-1 Relationship between first order biodegradation rates of test chemicals and the ratio of X_0 (initial putative competent degrader numbers) and C_0 (initial spiked chemical concentration)...... 129
Figure 7-2 Major catabolic genes (at the beginning and end of biodegradation assays) involved in the degradation pathways of test chemicals for different experimental conditions 131
Figure 7-3 Comparison of total bacteria counts in reactors with undiluted inocula during the biodegradation assays of different test chemicals at two different spiked chemical concentrations.. 132
Figure 7-4 Comparison of total bacteria counts in reactors with diluted inocula during the biodegradation assays of different test chemicals at two different spiked chemical concentrations....... 133
Figure 7-5 Monod no-growth curves, plotted for model predicted chemical concentration and the observed chemical concentration, for four chemicals (4-CP, 2,4-DCP, phenol and m-cresol) in diluted (10- fold) activated sludge inocula..... 135
Figure 7-6 Monod no-growth curves, plotted for model predicted chemical concentration and the observed chemical concentration, for four chemicals (4-CP, 2,4-DCP, phenol and m-cresol) in undiluted activated sludge inocula..... 136
Figure 7-7 Multidimensional scale plot comparing the bacterial communities based on genus level. The communities were taxonomically characterized at the beginning (i.e. t0) and end (i.e.t6) of biodegradation assays 137
Figure 7-8 Extended error bar plot considering the abundance profile of microbial domains in the 16S rRNA amplicon sequencing data for phenol degradation assay in diluted inoculum...... 139
Figure 7-9 Extended error bar plot considering the abundance profile of microbial domains in the 16S rRNA amplicon sequencing data for phenol degradation assay in undiluted inoculum. 140
Figure 7-10 Comparison of initial (t = initial) and final (t = final) absolute abundance of putative phenol degraders in the inocula used for phenol degradation assay at two different inocula concentration....... 141
Figure 7-11 Left: Absolute abundance (Log transformed) of genera that significantly changed over the duration of phenol biodegradation assays (both conditions) in rank order. 141

Figure 7-12 [A] Relationship between final and initial abundances for observed genera in phenol biodegradation assays with undiluted activated sludge inocula. [B] Volcano plot showing the 16S amplicon sequencing data at genus level (L6). 144
Figure 7-13 [A] Relationship between final and initial abundances for observed genera in phenol biodegradation assays with diluted (10-fold) activated sludge inocula. [B] Volcano plot showing the 16S amplicon sequencing data at genus level (L6).. 145
Figure 7-14 Comparison between initial (t_0) and final (t_{final}) absolute abundance of putative chemical degraders estimated with functional gene (with qPCR) and 16S rRNA amplicon sequencing (i.e. NGS) approach in biodegradation assays conducted with diluted inocula......147
Figure 7-15 Comparison between initial (t_0) and final (t_{end}) probable absolute abundance of putative chemical degraders estimated with functional gene (with qPCR) and 16S rRNA amplicon sequencing approach (i.e. NGS) in biodegradation assays conducted with undiluted inocula......147
Figure 7-16 Relationship between the experimentally determined biodegradation rates in grab activated sludge inocula and the biodegradation rates used in previously developed QSBR models for four test chemicals. 156
Figure 7-17 Relationship between the experimentally determined biodegradation rates in grab activated sludge inocula and the biodegradation rates used in previously developed QSBR models for four test chemicals at two different inocula concentration. 157

Acknowledgements

I would like to express my sincere gratitude to my supervisors Dr. Russell Davenport and Dr. David Werner for their support, patience, advice, motivation, immense knowledge and enthusiasm throughout this project. Their guidance helped me in all the time of research and writing of this thesis. I could not have imagined having a better supervisor and mentor for my PhD study. I would like to thanks my other secondary supervisors, Dr. Wojciech Mrozik and Dr. Paola Meynet for the productive discussion, assistance and guidance they offered to me in my laboratory work and thesis writing. Thank you to Marcos Quintela Baluja for his assistance in the design of initial experiments and for training in molecular techniques. I would also like to thank Dr. Jan Dolfing for giving out your wealth of experience, opinion and advice during the course of this work. I would like to thank Dr. Tomasz Puzyn and Maciej Barycki for providing me the training on QSARINs, particularly during my time at Laboratory of Environmental Chemometrics, University of Gdańsk, Poland. I would also like to thank members of staff within the School of Engineering; Paul Dunohoe, Sarah Smith, David Race, Amy Bell, Matthew Brown, Timothy Martin, Yongjie Yu, David Early, Sven Lahme and Dan Curtis for their technical support throughout my research. You guys are simply great. And to all other staff who contributed in one way or the other to the success of this project, to my colleagues and friends (Komo, Aidan, Amelie, Kaizer, Elias, David Walker, Cara, Sheldon, Fei, Magda, Aizat, Sanjeeb, Shamas, Pani, Vincenco, George, Evangelos, Tamara, Liana, Maria and Carolina), I say thank you all.

Finally, thank you to my family, particularly to my grand ma, mum, father and mother in laws, sister and my lovely wife Divya, for unconditional love, support and motivation throughout my thesis and my life in general.

This research project was funded by the *Engineering and Physical Sciences Research Council* (*EPSRC;* reference no: **EP/I025782/1**).

Chapter 1

General Introduction

Chapter 1. General Introduction

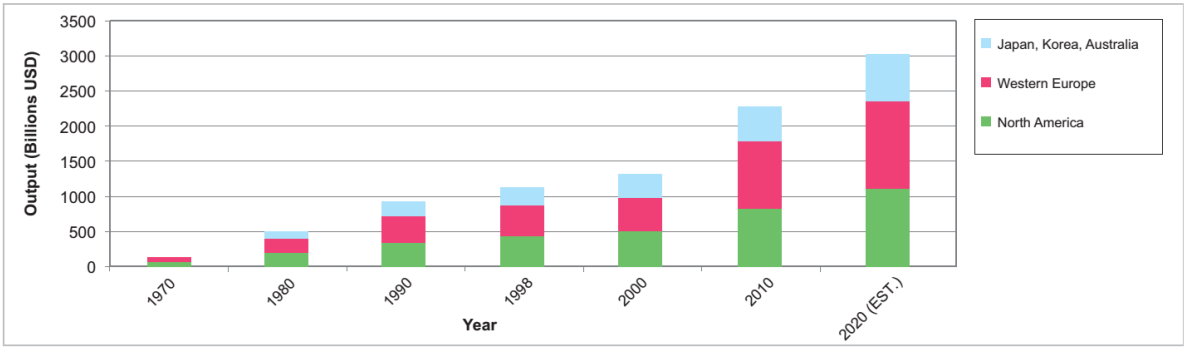
Over the last few decades, the reliance of our society on chemicals has increased [**Figure 1-1**] (UNEP, 2013). This has resulted in increased production, application and release of chemicals into the environment. As a consequence, chemicals pose a serious but uncertain threat to human health [**Table 1-1**] and the environment [**Table 1-2**] and the need for determining the chemical persistence, which is crucial for risk assessment and management of chemicals, becomes more pressing (ECHA, 2008).

Microbial biodegradation is an important process that determines the fate of anthropogenic chemicals in the environment (Grady, 1985; Martin, 2014; Kowalczyk *et al.*, 2015b). It can transform potentially harmful and toxic chemicals into less harmful products and can ultimately lead to their complete mineralization into carbon-dioxide, water and nutrients. Empirical laboratory biodegradation tests have thus become an indispensable step in the regulation of chemicals (Pavan and Worth, 2008; ECHA, 2012; Rücker and Kümmerer, 2012). However, current regulatory biodegradation testing methodologies used for the evaluation of biodegradation/persistence are expensive, time consuming and poorly reproducible (Kowalczyk *et al.*, 2015b; Martin *et al.*, 2017c). Furthermore factors affecting biodegradability are poorly understood and seldom studied (Kowalczyk *et al.*, 2015b). Therefore, an ability to reliably measure and predict biodegradation rates would help to accelerate and improve hazard and environmental risk assessment of chemicals, while reducing time, monetary cost and animal testing needed under current procedures.

Under the European Union (EU) REACH (Registration, Evaluation, Authorisation & restriction of CHemicals) Directive, the standard testing regime can be adapted by the use of non-test methods, such as Quantitative Structure-Activity Relationships [QSARs], if certain criteria are fulfilled (Nendza *et al.*, 2013; ECHA, 2016). Therefore, a predictive biodegradation model [i.e. Quantitative Structure Biodegradation Relationship (QSBR)] can be an alternative approach to standard tests to predict chemical persistence. Current biodegradation models, which are mostly based on the results derived from regulatory Ready Biodegradability Tests (RBTs), fundamentally underpin efforts to reliably predict chemical persistence. In such tests, biodegradation is viewed with respect to the intrinsic properties of chemicals, while measurement of the presence and activity of potential degraders in microbial communities receives little practical attention, but is known to be a key a variable factor in determining chemical persistence (Kowalczyk *et al.*, 2015b). Therefore, there is a need of techniques that determine not only the overall size of the inoculum, but also its quality (number of specific degraders) and the diversity of functional communities involved in the biodegradation process.

In this thesis, improvements to the existing approach for developing and verifying a predictive model and its reliability, respectively, will be proposed. In addition, the use of functional gene and 16S rRNA amplicon sequencing techniques towards better interpretation of biodegradation test data and prediction of chemical fate in the environment are studied.

A)



B)

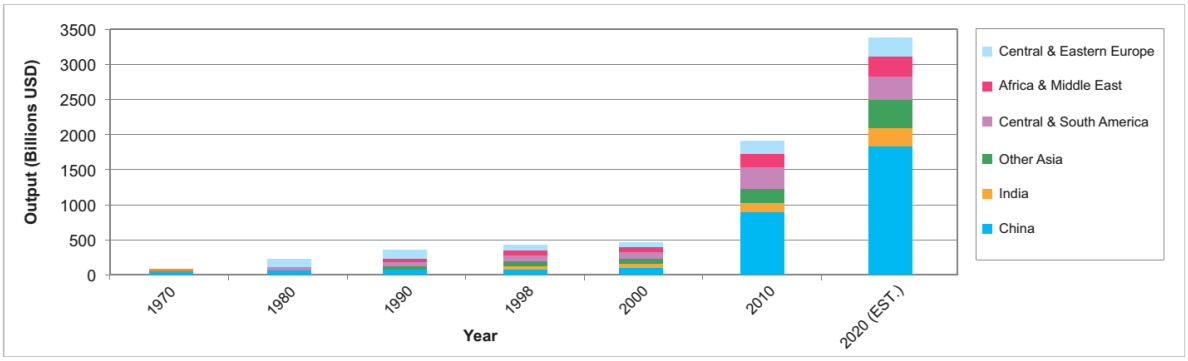


Figure 1-1 Chemical industry output for (A) Developed Regions and (B) Developing Regions & Countries with economies in transition s as categorized by UN Statistics Division. Data excludes pharmaceuticals. [Reproduced from (UNEP, 2013)]

Table 1-1 Overview of the environmental exposures associated with chronic non-communicable diseases

Disease	Expose during fetal development or early life	Exposure in later life	References
Asthma	Tobacco smoke, air pollution, bisphenol A	Tobacco smoke, air pollution, ecological exposure to Poly chlorinated biphenyls (PCBs)	(Midoro-Horiuti *et al.*, 2009; Barraza-Villarreal *et al.*, 2011; Donohue *et al.*, 2013; Hollams *et al.*, 2014)
COPD	Tobacco smoke, air pollution	Tobacco smoke, air pollution	(Upton *et al.*, 2004; Schikowski *et al.*, 2014)
Obesity	Tobacco smoke, bisphenol A, Persistent organic pollutants (POPs)	Bisphenol A, POPs	(Codru *et al.*, 2007; Cupul-Uicab *et al.*, 2012; Bhandari *et al.*, 2013)
Cancer	UV irradiation, arsenic, tobacco smoke	Air pollution, arsenic, POPs, many carcinogens	(Carpenter and Bushkin-Bedient, 2013; Norman *et al.*, 2014)
Hypertension	Organochlorine pesticides, tobacco smoke	Tobacco smoke, POPs, air pollution, arsenic	(Morley *et al.*, 1995; Peters *et al.*, 1997; Moon *et al.*, 2012; La Merrill *et al.*, 2013)

Table 1-2 Different chemicals related problems in the environment

Origin/usage	Selected chemical examples	Problems	References
Industrial chemicals (solvents, petrochemicals)	Tetrachloromethane, BTEX (Benzene, toluene, ethyl benzene, xylene)	Contamination of drinking water	(Schwarzenbach *et al.*, 2006)
Industrial products (additives, lubricants, flame retardants)	Phthalates, polychlorinated biphenyls, polybrominated diphenyl ethers	Biomagnification, long range transport	(Falkenmark and Rockström, 2004)
Consumer products (pharmaceuticals, hormones, personal-care products)	antibiotics, estrogens, ultraviolet filters	Antibiotics resistance, feminization of fish, multitude of (partially unknown) effects	(Lvovitch, 1973; Daughton and Ternes, 1999; Graham *et al.*, 2016)
Biocides (pesticides)	DDT, atrazine	Toxic effect and persistent metabolites, effects on primary producer	(Solomon *et al.*, 1996; Oki *et al.*, 2001)

Chapter2

Literature Review

Chapter 2. Literature Review

2.1 Human health and environmental implication of organic chemicals

Products of the chemical and pharmaceutical industries (such as personal care products, pharmaceuticals, fertilizers, pesticides, dyes, paints, food additives and preservatives, detergents) are essential contributors to the current standards of quality of life (Rücker and Kümmerer, 2012). On the other hand, they also pose a great threat to the environment, as many chemicals and pharmaceutical and their transformation products eventually end up in the environment, sometime after its intended use, even if they are used properly (Schwarzenbach *et al.*, 2006; Rücker and Kümmerer, 2012). Furthermore, various studies have highlighted that these chemicals can undergo long range transport via water, soil and the atmosphere following entry into the environment and can persist for much longer than predicted (Goldberg *et al.*, 1975; Kolpin *et al.*, 2002; Schwarzenbach *et al.*, 2006; Scheringer, 2009). More specifically, pharmaceuticals, hormones and other organic contaminants are identified in the environment and pose a significant human health threat and environmental issues (Heberer, 2002; Schwarzenbach *et al.*, 2006; Bellanger *et al.*, 2015; Hauser *et al.*, 2015; Sly *et al.*, 2016).

Various comprehensive reviews on the occurrence, fate and effects of chemicals on human health and the environment have been performed (Colborn *et al.*, 1993; Halling-Sørensen *et al.*, 1998; Vos *et al.*, 2000; Jones *et al.*, 2004; Fent *et al.*, 2006; Diamanti-Kandarakis *et al.*, 2009; Kümmerer, 2010; Bellanger *et al.*, 2015; Hauser *et al.*, 2015; Heath *et al.*, 2016; Sly *et al.*, 2016) and mostly highlight the prevalence of endocrine disruptors, chemotherapeutic drugs and antibiotics in the environment and the deleterious effects they can have on the exposed organism's health and development.

There is also a valid debate about what real effect these chemicals have on natural wildlife populations. For example, organisms exposed to endocrine-disrupting chemicals have shown different types of reproductive and developmental impairments; thyroid function in birds (Moccia *et al.*, 1986) and fish (Moccia *et al.*, 1981), and decreased fertility in birds (Shugart, 1980), fish (Leatherland, 1992), and mammals (Reijnders, 1986). Likewise, organisms and plants exposed to cytotoxic drugs have shown different types of deleterious effects on them; acute toxicity in zebrafish and adult fish (Kovács *et al.*, 2015), cytotoxic and genotoxic effect on zebrafish (Gajski *et al.*, 2016), *Daphnia magna* and *Ceridaphnia dubia (Kundi et al., 2016),* growth inhibition in green algae (*Pseudokirchneriella subcapitata* and the cyanobacterium *Synechococcus leopoliensis*) (Elersek *et al.*, 2016) and impairment in fertility of higher plants.

The impact of endocrine disruptors on human health has been thoroughly reviewed by Diamanti-Kandarakis *et al (Diamanti-Kandarakis et al., 2009)*. The authors revealed evidence for endocrine disruptors having effects on human reproduction, breast development and cancer, neuroendocrinology, thyroid, metabolism and obesity, cardiovascular endocrinology. Neu (1992) have highlighted the emerging threat of antibiotic resistance due to their voluminous and widespread synthesis, use and discharge into the environment (Neu, 1992). This resulted in rapid mobilisation of antibiotic resistant (AR) genes, and promoted the evolution of antibiotic resistance in bacterial strains not intrinsically resistant (Graham *et al.*, 2016). Furthermore, the environmental resistome (i.e. pool of AR genes) has been further altered due to the overuse of antibiotics and poor water quality in some parts of the globe, which can ultimately enhance the probability of AR acquisition in any exposed bacterial strain (Graham *et al.*, 2016). A study was conducted to review antibiotic use and their resistance in Europe (Goossens *et al.*, 2005). This study revealed that, the rate of antimicrobial resistance was higher in countries with higher antibiotic use. In addition, they also reported that the pattern of antibiotic resistance shifted from old-narrow spectrum antibiotic to new broad-spectrum antibiotics.

2.2 Fate of organic chemicals in the environment

Chemicals released into the environment may undergo a number of environmental fate and transport processes as shown in **Figure 2-1** (USGS, 2013). Chemicals in the environment may undergo various types of reactions that result in partial or complete transformation or degradation (i.e. mineralization to CO_2, H_2O, NO_3^-, SO_4^{2-}, and other inorganics), or undergo no change at all. Abiotic mechanisms like photolysis, hydrolysis, sorption, oxidation and volatilization, as well as microbial process such as biodegradation, result in removal of chemicals from natural or engineered ecosystems, but may also lead to formation of long-lived transformation products, unless transformation results in full degradation (i.e. mineralization).

Chemicals (and transformation products) which tend to persist in the environment are targeted for risk management measures (Pavan and Worth, 2008; Rücker and Kümmerer, 2012). These groups of chemicals are of global environmental concern, as they are not only persistent, but offer the chance of bioaccumulation in environmental biota, and they may also be transported into other environmental compartments far from their original source (Blais, 2005).

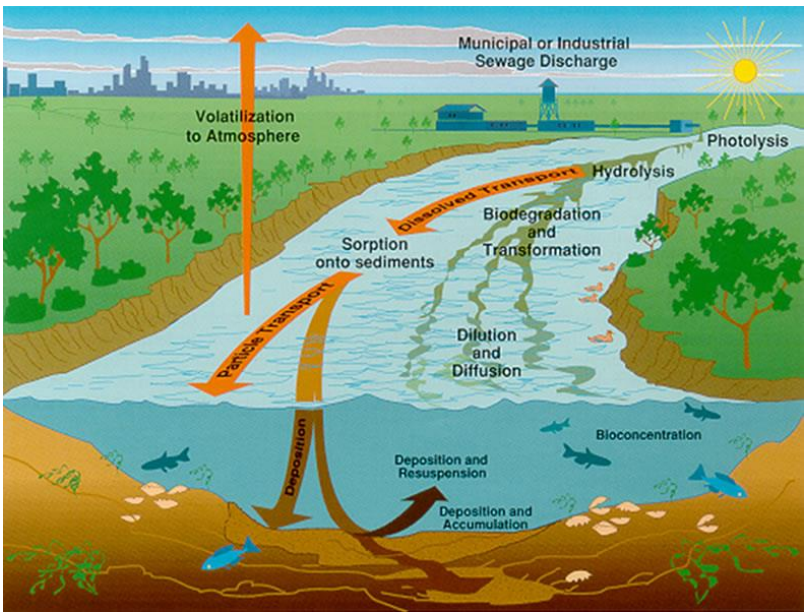


Figure 2-1 Transport and fate of chemicals entering the aquatic environment [Reproduced from: (USGS, 2013)]

2.3 Chemical persistence and its assessment

The persistence of a chemical is defined as the time the chemical remains in a particular environment before it is physically transported to another environmental compartment, and/or is chemically or biologically degraded (Pavan and Worth, 2008). A chemical's half-life (50% degradation) in an environmental compartment is used to evaluate the chemical's persistence (i.e. a chemical's half-life is used to assess the environmental fate of the chemical and the risk it presents) (Pavan and Worth, 2008; Bu *et al.*, 2016; Martin *et al.*, 2017b). As mentioned earlier, poor degradation (long half-life), or persistence (very-long half-life), cause serious problems to human health and the environment, which includes potential accumulation of chemicals in organisms and the food chain, and ultimately may result in toxic effects on organisms (Sumpter, 2009). In the European Union (EU), persistent chemicals will be regulated under REACH, and will be classified as Substance of Very High Concern (SVHC) [e.g., Trichloroethylene, Bis(2-ethylhexyl) phthalate (DEHP), and Diisobutyl phthalate (DIBP)] and banned from the market or use after a given date unless an authorization is granted for their specific use (EC, 2006). Furthermore, the Water Framework Directive (WFD) [applicable in European Union], in 2000, identified 33 chemicals classified as priority substances (20 chemicals) or priority hazardous substances (13 chemicals), which are subject to control measures over the subsequent 20 years, and recently included an additional 13 substances in the revised list (EC, 2008). Environmental Quality Standards (EQSs) for these chemicals will be used to set discharge permits to water bodies, so that chemical emissions do not exceed EQS limits within the receiving water (EC, 2011). The protection of public health and the environment from such priority substances thus demands

energy intensive and expensive treatment facilities and systems for effluent treatment and drinking water production (Eggen *et al.*, 2014; Stamm *et al.*, 2015). On the other hand, chemicals (and their transformation products) with short half-lives (ready degradation) will avoid these problems and are considered safer for both the environment and human health. Regulatory bodies across the world have set different persistence criteria with the common theme of protecting human health and the environment. **Table 2-1** summarizes the criteria for the identification of persistent chemicals under different regulatory programmes.

Table 2-1 The persistency criteria set by different regulatory bodies [Adapted from ; (Pavan and Worth, 2008)]

Regulatory programme	**Criteria**
European Union (EU); identification of PBTs	Half-life> 60 days in marine water, or > 40 days in fresh or estuarine water or, >180 days in marine sediment, or >120 days in fresh water sediment, or > 120 days in soil
European Union (EU); identification of vPvBs	Half-life> 60 days in marine water, fresh or estuarine water or, >180 days in marine, fresh or estuarine sediment, or > 180 days in soil
US EPA	Transformation half-life > 2 months
Canada (Canadian Domestic Substances list)	Half-life in air > 2 days, water > 6 months, sediments>1 year, soil > 6 months

PBTs: Persistent, Bio-accumulative and Toxic

vPvB: very Persistent, very Bio-accumulative

The persistence of a chemical is determined empirically by conducting a degradation experiment using an experimental test system – a chemical that is degraded in that test system is not considered persistent in an environment (Pavan and Worth, 2008). The extent of degradation is used to characterize the degradation process and can be primary or ultimate degradation.

Primary degradation: Conversion of parent chemicals to other derivatives, which have their own properties and fate.

Ultimate degradation (Mineralization): Complete degradation of chemicals to stable inorganic species as mentioned above.

Similarly, the degradation can also be characterised by the nature of the process involved for degradation, such as:

Abiotic degradation: Degradation of chemicals by reactions like photolysis, oxidation, reduction and hydrolysis.

Microbial degradation (Biodegradation): Degradation of chemicals brought about by enzymatic reactions in microorganisms.

In natural and engineered eco-systems, microbial degradation is the most important process for removal of chemicals from the environment and can result in complete degradation of chemicals (Grady, 1985; Rücker and Kümmerer, 2012). The extent of biodegradation depends on different factors (e.g., physicochemical properties of chemical, activity of degrading microorganisms, environmental conditions, etc.) (Itrich *et al.*, 2015). A chemical may be readily biodegradable in one site, but not in another site due to the different biodegradation capacity of that site. This is mostly attributed to the characteristics of microbial inoculum, as presence or absence of competent degraders in the inocula will determine both the probability and extent of biodegradation (Goodhead *et al.*, 2014; Itrich *et al.*, 2015; Kowalczyk *et al.*, 2015b; Martin *et al.*, 2017c).

2.3.1 Biodegradation Kinetics

Kinetic data form the basis for predictive models (see below), which would enable determination and prioritization of chemical persistence, thereby potentially reducing laboratory testing of chemicals and gaining a greater understanding of those factors that lead to persistence (Battersby, 1990). Such an approach could lead to the targeted manufacture of chemicals that are benign by design – so-called green chemistry (Rücker and Kümmerer, 2012). Commonly observed degradation kinetics are first order, pseudo-first order, second order and zero order (Simkins and Alexander, 1984; Battersby, 1990). In first order kinetics, the biodegradation rate constant is proportional to a single factor, typically the concentration of the chemical. This is the most commonly used type of kinetics to describe the biodegradation of organic chemicals in natural or engineered ecosystems. However, in such systems, there may be one or a number of different mechanisms occurring at the same time and, thus first order kinetics might not be a true representation of the biodegradation process (Battersby, 1990). For example, bacteria might play a role in biodegradation of chemicals, which leads to pseudo-first order rates, where second-order reaction occurs, but apparently exhibit first order rates. Likewise, the biodegradation rate can be a function of both chemical concentration and bacterial population and would then exhibit second order kinetics. The kinetics of biodegradation is commonly measured with a half -life (i.e. $t_{1/2}$). If the biodegradation kinetics follows first order, the biodegradation rate constant (k) is related with the half-life through Equation 1.

$$t_{1/2} = \frac{ln2}{k} \quad [1]$$

The Monod equation (Equation 2) is another widely used non-linear mathematical model to represent the bacterial growth kinetics (Monod, 1949), which relates the specific growth rate of bacteria to the concentration of a rate limiting substrate as follows:

$$\mu = \mu_{max} \frac{S}{K_s + S} \quad [2]$$

Where μ = specific growth rate (s^{-1}), μ_{max} = maximum specific growth rate (s^{-1}), S = substrate concentration (moles C m^{-3}), K_s = substrate saturation constant (moles C m^{-3}) [i.e. substrate concentration at half μ_{max}].

The specific growth rate (μ) reaches its maximum (i.e. maximum specific growth rate [μ_{max}]), when another parameter becomes the limiting factor towards microbial growth. The growth of bacteria occurs through substrate utilization and reaches its maximum at high substrate concentration. This scenario is mostly valid for laboratory biodegradation studies, where the test chemical serves as a sole carbon and energy source for microorganisms, but may not be valid in natural and engineered eco-systems where alternative carbon and energy sources are present (Battersby, 1990). The Monod biodegradation kinetics of microbial growth depends on the initial chemical concentration. At low chemical concentration, the degradation kinetic follow first order; which becomes second order with increasing substrate concentration due to microbial growth; and finally, zero-order, where microbial growth remains constant with an increasing substrate concentration (**Figure 2-2**). Different factors should be considered when selecting the appropriate model to describe the biodegradation data, which include; (1) number of substrates applied; single substrate or mixture of substrates, (2) types of inocula used; single bacterium or mixed bacterial inocula and (3) limiting factors that affect the degradation kinetics, such as chemical and biomass concentration and the presence or absence of other nutrients required for the growth of microorganisms (Simkins and Alexander, 1984).

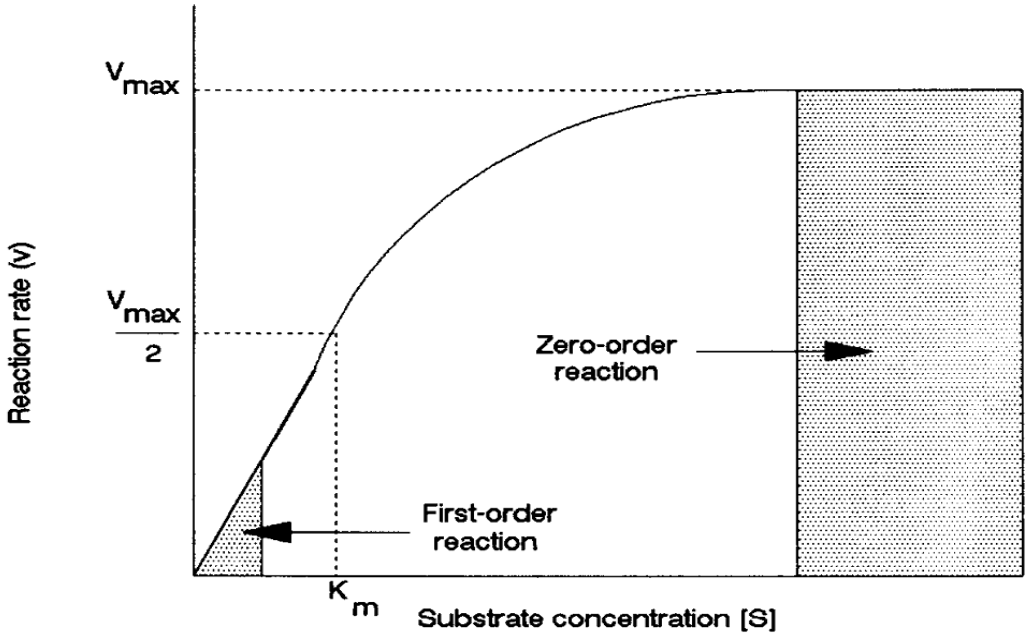


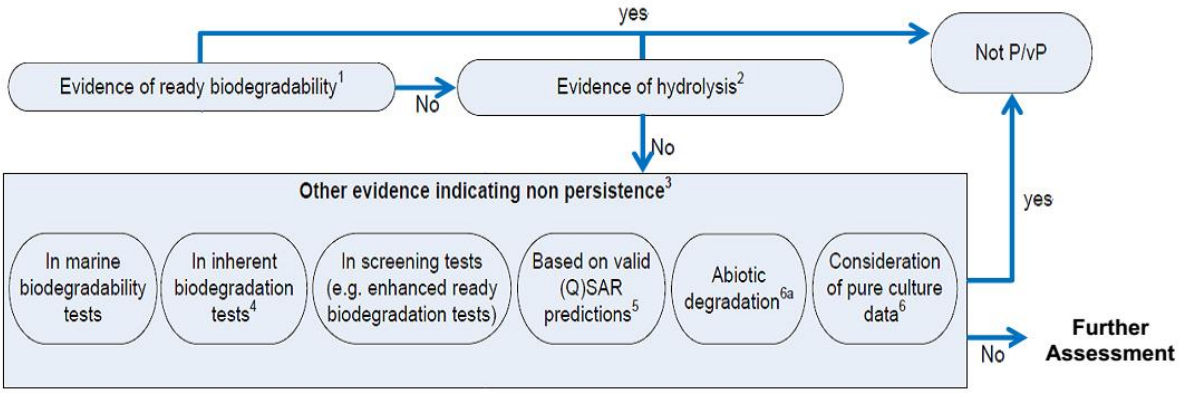
Figure 2-2 Effect of substrate concentration on kinetics of reaction [Reproduced from : (Battersby, 1990)]

2.4 Biodegradation, biodegradation tests and risk assessment

Natural and engineered ecosystems are subject to a host of chemicals, both natural and synthetic. As mentioned earlier, biodegradation is the principal fate process for the removal of chemicals from these systems, and the ability to predict biodegradation thus plays an important role in determining eventual environmental concentrations, exposure risks and ultimately the long-term effects of chemicals. Biodegradation tests have thus become an indispensable step in the regulation and risk assessment of chemicals (Kowalczyk *et al.*, 2015b); determining their classification and labelling [**Figure 2-4**], hazard assessment (such as Persistence, Bioaccumulation and Toxicity) and exposure assessment [**Figure 2-5**], with a regulatory shift towards identifying persistent chemicals (Pavan and Worth, 2008; ECHA, 2012; Rücker and Kümmerer, 2012).

A range of biodegradation test methods with pure and mixed cultures have been developed in laboratories to study the fate of chemicals, with most studies focusing on evaluating the fate of chemicals in aquatic systems, especially, waste water treatment processes (Reuschenbach *et al.*, 2003). The Organisation for Economic Cooperation and Development (OECD) proposed the guidelines for testing the biodegradability of chemicals (OECD, 1994), and can be divided into three groups: (1) ready biodegradability tests (a screening test), (2) inherent biodegradability test (also a screening test) and (3) simulation test (OECD, 2003). These tests are an important part of regulatory frameworks and help to assess the likelihood of chemical persistence in the environment. A strategy for degradation assessment and testing in the

context of PBT/vPvB [persistence, bioaccumulation and toxicity / very persistence and very bio accumulative] assessment was stipulated under REACH in **Figure 2-3** (ECHA, 2008).



[1]If the substance fulfils the criteria for ready biodegradability, there is no reason to perform further biodegradation tests for the PBT/vPvB assessment.
[2]The mass balance should be considered if it needed to address the concern for losses by volatilization and absorption to glassware. Rapid hydrolysis also needs to be shown across all environmentally relevant pH.
[3]All available information on biodegradation, including testing, non-testing and monitoring data, should be considered and useful for weight of evidence approach
[4]Results of a Zahn-Wellens test (OECD TG 302B) or MITI II test (OECD TG 302C) only (not SCAS-test) may be used to confirm that the substance does not fulfil the criteria for P provided that certain additional conditions are fulfilled.
[5]Consider both the validation status of any QSAR model and whether the substance for which predictions are made may be regarded as being within the applicability domain of the model
[6]The biodegradation data obtained with pure culture(s) studied; single species or mixture of species, cannot be used on their own within persistence assessment but should be considered as part of a *Weight-of-Evidence* approach
[6a]Data derived from abiotic studies (e.g. photodegradation, oxidation, reduction) cannot be used on their own within persistence assessment, but may be used as part of a *Weight-of-Evidence* approach
Figure 2-3 The first step in the persistence assessment of chemical under REACH [Reproduced from; (Martin, 2014)] P/vP = persistent/very Persistent

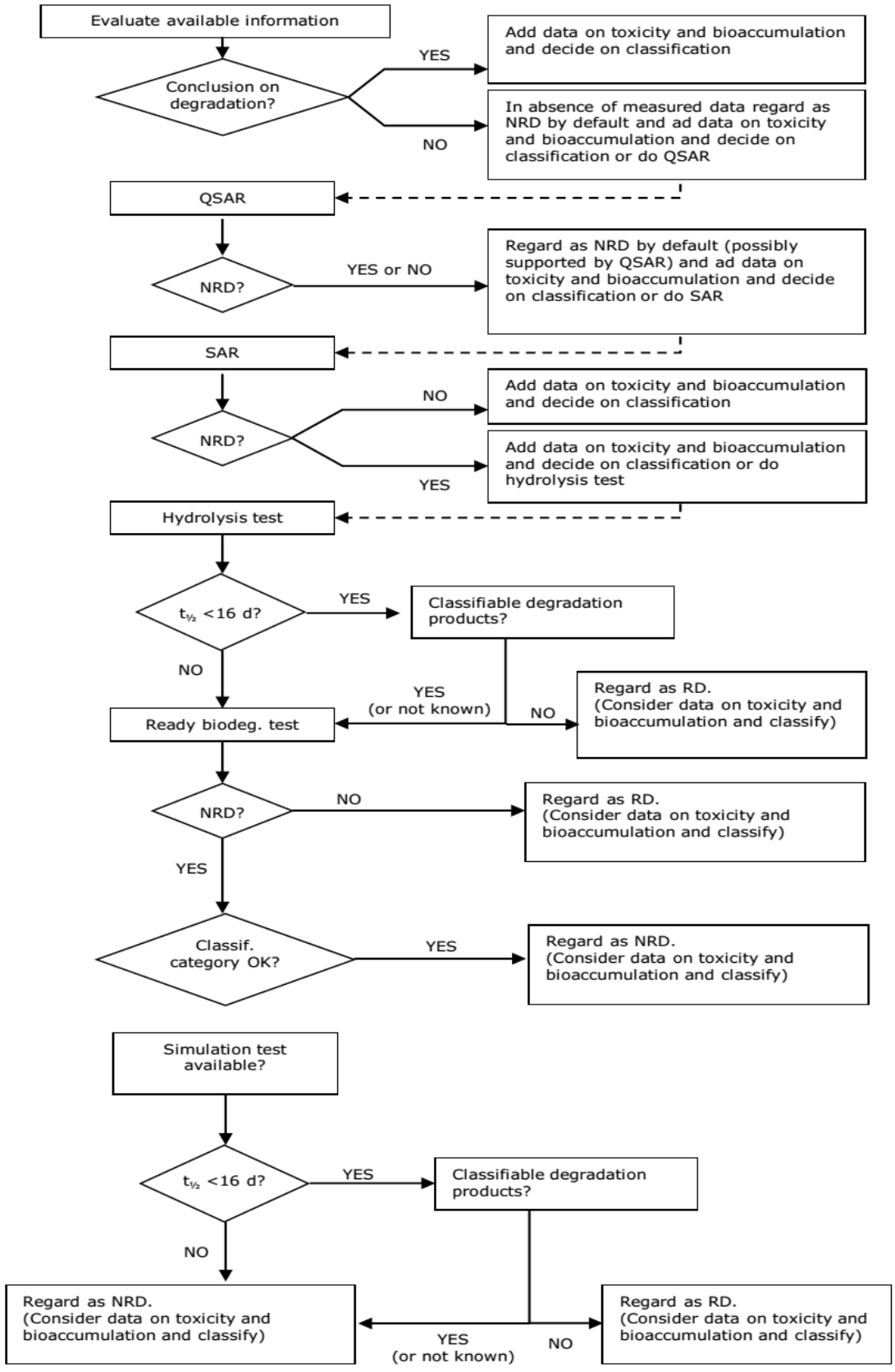


Figure 2-4 Schematic overview of using degradation data in classification and Labelling. RD and NRD refers to ready biodegradability and non-ready biodegradability, respectively. [Adapted from;(ECHA, 2012)]

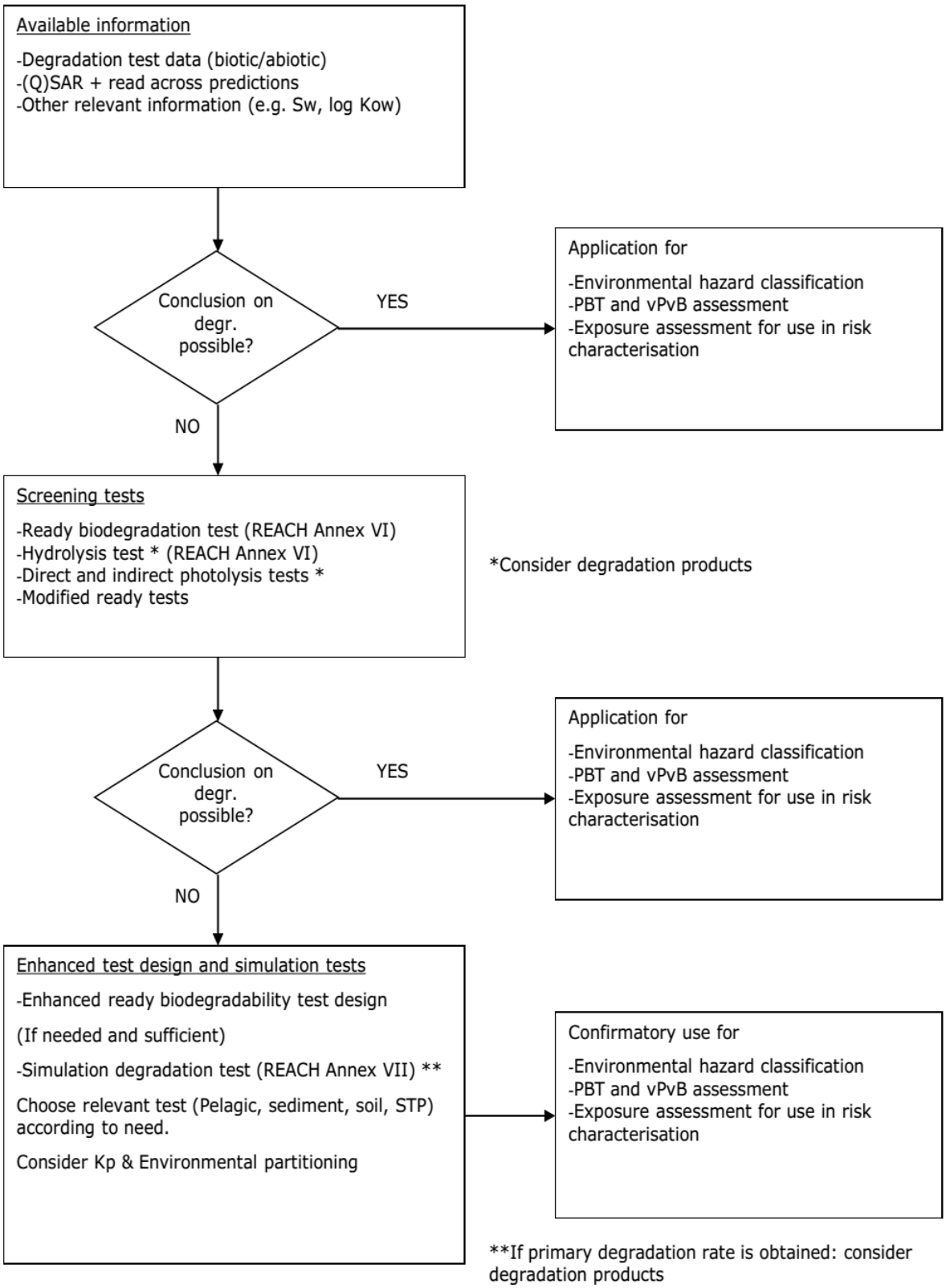


Figure 2-5 Summary of decision scheme on degradation for three different regulatory needs (i.e. Environmental hazard classification, PBT/vPvB assessment and Exposure assessment for use in risk characterization.[Adapted from;(ECHA, 2012)]

Under REACH, the selection of an appropriate biodegradation test depends on the production level and properties of chemicals (**Table 2-2**) (ECHA, 2012). Ready biodegradability tests (RBTs) are the first-tier stringent screening tests conducted under aerobic conditions to

identify if a chemical is rapidly biodegradable or not. The results of these methods are based on the measurement of dissolved organic carbon (DOC) [OECD 301 A, OECD 301 E], CO_2 production [OECD 301 E], the determination of biological oxygen demand (BOD) [OECD 301 C, OECD 301 D, OECD 301 F] and the quantification of inorganic carbon [OECD 310] (Kowalczyk *et al.*, 2015b). These tests use small amounts of inoculum from domestic sewage, activated sludge, secondary effluent, surface water or soil, and should not be artificially pre-adapted to the test chemical through previous exposure to either the test chemical or structurally related chemicals. In addition, the test chemical is provided as the sole source of carbon for energy and growth. A positive result in these tests can be considered indicative of rapid and ultimate degradation in most environments including biological sewage treatment plants (ECHA, 2012).

Second tier tests (i.e. inherent biodegradability tests) are required, when a chemical fails the screening test. The inherent biodegradability of a test chemical is measured by specific analysis (primary biodegradation) or by non-specific analysis (ultimate biodegradation). Test for inherent biodegradability includes, Modified SCAS (Semi-continuous activated sludge) Test (OECD, 1981a), Zahn-Wellens Test (OECD, 1992b) and Modified MITI Test (II) (OECD, 1981b). All these tests are conducted under aerobic conditions with higher inoculum concentration, and the biodegradation rate and/ or extent are measured. The test procedure offers a greater chance of detecting biodegradation compared to screening tests and therefore, if an inherent test is negative, this could be indicative for potential environmental persistence (ECHA, 2012).

The higher tier test (i.e. simulation tests) can be conducted under both aerobic and anaerobic conditions, and aims to assess the rate and extent of biodegradation in a laboratory system designed to represent specified environmentally relevant conditions (i.e. aerobic or anaerobic treatment stage of a wastewater treatment plant (WWTP) or environmental compartment, such as fresh or marine surface water, soils, and sediments) (ECHA, 2012).

2.4.1 Limitations of OECD tests

The OECD tests, especially screening tests, suffer a number of well documented limitations (Kowalczyk *et al.*, 2015b), which include, (1) a majority of tests are conducted under environmentally unrealistic conditions (e.g., use of high test chemicals concentration that are unlikely to occur in environment), (2) tests are conducted under standardized incubation conditions, which doesn't reflect highly variable environmental conditions, (3) variation in the inoculum test volume from test to test and lack of consideration of inoculum quality or quantity and protocol regarding inoculum preparation, and (4) lack of consistency in test pass level and test durations

Table 2-2 Required test data of interest for the integrated testing strategy (ITS) on degradation

Tonnage band (t/y/registrant)	Required degradation data	Other relevant information
1-10	Ready biodegradability[5]	
10-100	Ready biodegradability Hydrolysis	Log KOW Vapour pressure Water solubility Adsorption/desorption
100-1000	Ready biodegradability Hydrolysis Simulation of biodegradability in water[1] Simulation of biodegradability in sediment[2] Simulation of biodegradability in soil[3]	Log KOW Vapour pressure Water solubility Adsorption/desorption Dissociation constant Degradation products BCF[4]
>1000	Ready biodegradability Hydrolysis Simulation of biodegradability in water[1] Simulation of biodegradability in sediment[2] Simulation of biodegradability in soil[3] Further testing shall be proposed if the CSA indicates a need for additional data on the degradation of the substance	Log KOW Vapour pressure Water solubility Adsorption/desorption Dissociation constant Degradation products BCF[4]

1. Not needed if the substance is highly insoluble in water and/or is readily biodegradable

2. Not needed if the substance is readily biodegradable and/or direct and indirect exposure of sediment is unlikely

3. Not needed if the substance is readily biodegradable and/or direct and indirect exposure of soil is unlikely

4. Not needed if the substance has a low potential for bioaccumulation (for instance a log Kow < 3) and/or a low potential to cross biological membranes and/or direct and indirect exposure of the aquatic compartment is unlikely.

5. The study does not need to be conducted if the substance is inorganic

2.5 Predictive modelling, an alternative to biodegradation test

Current laboratory testing methodologies used for the evaluation of biodegradation/persistence are expensive, time consuming and poorly reproducible (Goodhead *et al.*, 2014; Kowalczyk *et al.*, 2015b). To date only 4214 of the estimated 120,000 chemicals (15,837 currently registered under EU REACH legislation; ECHA, 2017, Accessed Date:23/06/2017) (ECHA, 2017) have been reliably screened for their biodegradability (OECD, 2017). Performing such tests for the remaining existing and new chemicals is a laborious, costly and perhaps unachievable task. Furthermore factors affecting

biodegradability are poorly understood and seldom studied (Rücker and Kümmerer, 2012; Kowalczyk *et al.*, 2015b). Therefore, development of predictive biodegradation models and their application towards predicting the biodegradation rates would help to accelerate and improve environmental risk assessment of chemicals (Pavan and Worth, 2008; Rücker and Kümmerer, 2012; Martin *et al.*, 2017a).

2.5.1 Quantitative Structure Activity Relationships (QSARs)

Biodegradation models are broadly classified into structure-activity relationship (SAR) and quantitative structure-activity relationship (QSAR) models – collectively referred to as (Q)SARs. *A QSAR is a theoretical mathematical model that can be used to quantitatively or qualitatively predict the physicochemical, biological (e.g. an (eco) toxicological endpoint) and environmental fate properties of compounds from the knowledge of their chemical structure* (ECHA, 2016). QSARs are widely used in diverse fields (e.g., toxicology, environmental science, electrochemistry, chemistry, biomedical science) (Okey and Stensel, 1996; Roy *et al.*, 2011; Lee and von Gunten, 2012; Mikolajczyk *et al.*, 2015). In recent years, the application of QSARs to biodegradation (Quantitative Structure Biodegradability Relationship, QSBR) has been advocated (Pavan and Worth, 2008; Rücker and Kümmerer, 2012), providing a valuable prediction of relative or absolute biodegradation rate and/or transformation products of biodegradable chemicals strictly on the basis of chemical structures without having to undertake laboratory testing. Hence, such correlations could provide a powerful tool to forecast the environmental fate of chemicals and assist in ranking them for in-depth evaluation. However, current QSBRs suffer a number of limitations that restrict their use in biodegradation assessments (see below)

Molecular Parameters (Descriptors)

The rate at which a compound is degraded by the microorganism depends on a number of intrinsic factors that affect the mechanism of enzyme reactions such as (i) the reaction of electrophilic groups of the substrate with nucleophilic groups of the enzyme centre and vice versa, (ii) size and spatial arrangement of the molecule which affects enzyme substrate binding, and (iii) adsorption and transport of a compound across the cell membrane. Therefore, physicochemical properties of electronic, steric and lipophilic nature can be used to express the biodegradability of a given compound (Pitter and Chudoba, 1990).

Different molecular parameters (descriptors) can be applied to describe the above-mentioned properties. Thus, descriptors act as an important tool in QSBR modelling as they encode the numerical information on different chemical features of compounds (chemical structure, physicochemical properties)(Todeschini and Consonni, 2008).

The most commonly used molecular descriptors that describe electronic, steric and lipophilic properties of compounds in QSBR models are the Hammett constant, 1-Octanol Water Partitioning coefficient (Banerjee *et al.*, 1984), Van der Waal's radii (Paris *et al.*, 1982; Paris *et al.*, 1983; Paris and Wolfe, 1987), molecular refractivity, standard Gibbs free energy (Yang *et al.*, 2006b), molecular weight, structural/molecular fragments (Loonen *et al.*, 1999; Tunkel *et al.*, 2000) and connectivity indices (Okey and Stensel, 1996).

Modelling Biodegradation

Qualitative models predict if a chemical is biodegradable or not, using structural fragments or other properties of a chemical as a molecular descriptor (Gamberger *et al.*, 1996a; Gamberger *et al.*, 1996b; Loonen *et al.*, 1999; Tunkel *et al.*, 2000). On the other hand, quantitative models predict the extent and sometimes biodegradation pathways of chemicals, mostly using different structural and physicochemical properties of a chemical (such as molecular descriptors). **Table 2-5** summaries some of the developed qualitative and quantitative biodegradation models. Development of QSBRs has been relatively slow as compared to the proliferation of QSARs for other purposes. Most of the QSBR models were developed from a small set of homologous chemicals and have limited applicability (Sabljic and Peijnenburg, 2001; Pavan and Worth, 2008). Lack of consistent experimental biodegradation data and unavailability of standardized and uniform biodegradation databases are considered as the major obstacle that precluded the development of reliable QSBR models. In the late 1990s, two databases for biodegradation became available [(i) the BIODEG database and (ii) the MITI database] which ultimately resulted in the development of better qualitative biodegradability models. In addition, important advances were made in: (a) formulation of qualitative biodegradability rules by the application of artificial intelligence and (b) development of reliable qualitative biodegradability models by application of partial least squares (PLS) discriminant analysis and (c) utilization of MultiCASE programming tools to select the appropriate structural fragments that affects the biodegradation of organic pollutants (Raymond *et al.*, 2001; Sabljic and Peijnenburg, 2001).

Biodegradation Data and Database

Several biodegradation databases are available for direct evaluation of biodegradability and qualitative model development. Some commonly used biodegradation databases and their features are mentioned in **Table 2-3** below. The biodegradation data of chemicals that are presented in the BIODEG and MITI databases are the results of Ready Biodegradability Tests (RBTs) or screening tests (Martin *et al.*, 2017c). In screening tests; chemical, environmental inoculum and mineral media are combined together to assess the biodegradability of the chemical at different conditions based on a pass/fail basis and were not intended for

measuring rates. These biodegradation results are therefore limited in their ability to assess biodegradation kinetics and thus develop predictive models to accurately predict degradation kinetics.

Table 2-3 List of the most commonly used databases for biodegradation studies

Database	Information	No. of Compounds covered	References
BIODEG	Biodegradation of chemicals in several types of experiments (biological treatment simulation, screening tests, field studies, etc.) under variety of experimental conditions (aerobic, anaerobic)	815	(Rücker and Kümmerer, 2012) (Sabljic and Peijnenburg, 2001; Pavan and Worth, 2008)
MITI-I	Screening test result for ready biodegradability in an aqueous medium and is described in OECD and EU test guidelines	900	(Sabljic and Peijnenburg, 2001; Pavan and Worth, 2008; Rücker and Kümmerer, 2012)
UM-BBD	Microbial bio-catalytic reaction and biodegradation pathways for chemical compounds	1200	(Gao *et al.*, 2010)

2.5.2 Different Approaches for Modelling Biodegradation

Simple Regression Approaches

The QSBRs models developed during the 1980s and 1990s were statistical correlations between biodegradability endpoints and molecular descriptors (Banerjee *et al.*, 1984; Paris and Wolfe, 1987; Okey and Stensel, 1996). Most of the developed models were class specific models; as significant and mechanistically reasonable correlations between biodegradability and molecular structure were established between the congeneric series of chemicals. For example, models have been developed to predict the biodegradation of a limited number of alcohols (Yonezawa and Urushigawa, 1979), chlorophenols and chloroanisol, (Banerjee *et al.*, 1984), para-substituted phenols (Paris *et al.*, 1983) and meta-substituted anilines (Paris and Wolfe, 1987). Most of the developed QSBR models used octanol-water partition coefficient (K_{ow}), Van der Waal's radii, molecular connectivity indices, and the Hammett constant as

molecular descriptors. These descriptors mostly reflect molecular properties of the molecules rather than individual fragment contributions. Basically, there is a good correlation between the physicochemical properties or molecular descriptors and the biodegradation rates (Pavan and Worth, 2008; Lee and von Gunten, 2012). Despite this, these models have not been used for regulatory and risk assessment purposes. The applicability of class specific models is generally limited to predict the biodegradation of chemicals inside the co-generic classes.

Group contribution Approaches

In recent years, several new and better qualitative and quantitative biodegradation models have been developed using advanced computational and statistical techniques. Biodegradation models based on group contribution approaches are the most widely accepted approach in modelling (Pavan and Worth, 2008). This approach was mostly developed to expand the applicability of QSBR to large and structurally diverse sets of chemicals. In this approach, a number of fragments are obtained from the decomposition of the chemical structure, and the model expresses biodegradability as a function of the contribution of each fragment in the molecules. The weighted molecular fragments are the model descriptors. It makes an assumption that the molecular fragments can have either an enhancing effect or a retarding effect on biodegradability. Each molecule from the training set is first decomposed into the fragments, the fragments' contribution weights are assigned and its biodegradability is evaluated based on the weights of the fragments. The weight of the individual fragments can be assigned using different statistical techniques (examples include: linear and non-linear regression modelling, partial least squares (PLS), and neural networks). The endpoint biodegradability can be either semi-quantitative rates expressed in days, weeks and month or Boolean (yes/no) determination of ready biodegradability (Raymond *et al.*, 2001; Sabljic and Peijnenburg, 2001; Pavan and Worth, 2008).

Biodegradation models developed from a group contribution approach allows a structurally diverse set of chemicals to be analyzed. However, this approach is highly dependent on the type and number of selected fragments and the way a molecule is fragmented (Pavan and Worth, 2008). To avoid this limitation, a MultiCASE software has been developed (Klopman *et al.*, 1994). With this software, it is possible to generate all possible fragments of the molecules and subsequently select the statistically most significant fragments for the endpoint of interest. The fragments are then used to develop the regression model between the selected fragments and the endpoint. With the MultiCASE approach, the fragments that enhance the degradation of the molecules are termed as Biophores, whereas those that inhibit the degradation are called Biophobes (Sabljic and Peijnenburg, 2001; Pavan and Worth, 2008).

Examples of qualitative biodegradation models based on the group contribution approach that use different sets of structural fragments have been published (Loonen *et al.*, 1999; Tunkel *et al.*, 2000). In addition, the Biodegradation Probability Program (BIOWIN) developed by SRC on behalf of EPA is also based on the group contribution approach (Pavan and Worth, 2008). This software incorporates six different models (described in **Table 2-4**) that estimate the probability of rapid biodegradation of organic compounds in the presence of a mixed community of environmental microorganisms.

Table 2-4 Six different Biowin models included in Biodegradation Probability Program (BIODEG) [Adapted from; (Pavan and Worth, 2008)]

Model	Model types	Biodegradation data	Nature of prediction
Biowin1	Linear probability model	BIODEG	Quantitative prediction of the probability that a chemical biodegrades fast
Biowin2	Non-linear probability model	BIODEG	Quantitative prediction of the probability that a chemical biodegrades fast
Biowin3	Ultimate biodegradation model	Expert survey	Semi-quantitative prediction based on the scoring of persistence; 5 = hours, 4 = days, 3 = weeks, 2 = months, 1=longer
Biowin4	Primary biodegradation model	Expert survey	Semi-quantitative prediction based on the scoring of persistence; 5 = hours, 4 = days, 3 = weeks, 2 = months, 1=longer
Biowin5	Linear probability model	Japanese MITI	Quantitative prediction of the probability that a chemical is readily biodegradable or non-readily biodegradable
Biowin6	Non-linear probability model	Japanese MITI	Quantitative prediction of the probability that a chemical is readily biodegradable or non-readily biodegradable

MITI: Ministry of International Trade and Industry

Expert System

Consideration of pathway information in modelling the biodegradability of chemicals is a new concept and this led to the development of QSBR studies mainly based on expert systems (i.e. artificial intelligence approach) (Baker *et al.*, 2004; Rücker and Kümmerer, 2012). In expert systems, prediction is generated on the basis of rules developed by experts and sometimes facilitated with external databases. An expert system can be either; (i) a 'knowledge based' system, where rules and hierarchy of biodegradation pathways are summarized by expert knowledge or (ii) a 'machine learning or inductive' system, where biodegradation rules and a hierarchy of rules are developed without human input (Pavan and Worth, 2008). Gramberger et al. (Gamberger *et al.*, 1996a; Gamberger *et al.*, 1996b) used inductive machine learning techniques to develop the rules based on structural requirements for slow and fast biodegradation.

Generally, the expert system is qualitative in nature, but they can be combined with other models to provide quantitative assessments (Pavan and Worth, 2008). META / MultiCASE is a hybrid approach which makes combined use of META and MultiCASE (Klopman *et al.*, 1994). META is an automatic rule induction program which helps in qualitative prediction of the aerobic biodegradation pathways. META and MultiCASE programs are used in conjugation to evaluate the fate of disposed chemicals by estimating the biodegradability and the nature of biodegradation products under conditions that may model the environment. In this approach, MultiCASE is used to identify the structural fragments that inhibit the biodegradation; these fragments are included in the dictionary, such that META can exclude the compounds that contain inhibiting fragments before applying any transformation rule (Pavan and Worth, 2008).

2.5.3 Conditions for using QSAR results

The acceptance of QSBR model predictions depend on their reliability and relevance (Nendza *et al.*, 2013). According to Annex XI of Registration, Evaluation, Authorisation & Restriction of Chemicals (REACH), the use of QSAR model for regulatory purposes is valid if: (i) the model is developed in accordance to OECD principles, (ii) the evaluated substance is within the applicability domain of the model, (iii) the predicted result is suitable to use for regulatory purposes and (iv) adequate documentation of the method is provided (Nendza *et al.*, 2013; ECHA, 2017).

The rigorous validation of quantitative models such as QSBR requires both a defined endpoint and applicability domain, should be associated with appropriate measures of goodness of fit, robustness and predictability, and should be expressed in the form of an unambiguous algorithm capable of providing a potential mechanistic interpretation of

biodegradation if possible, as per OECD principles for QSAR validation (Netzeva *et al.*, 2005).

Table 2-5 Examples of different QSAR models. The nature of QSBR model (Qualitative or Quantitative) are indicated, different kinds of independent variables used by the models are mentioned with techniques used for modelling

References	Model Prediction	Descriptors Used	Modelling Techniques
Tunkel *et al.* (Tunkel *et al.*, 2000)	Qualitative	MW, Structural fragments/group contributions	Multiple linear and nonlinear regression
Loonen *et al.* (Loonen *et al.*, 1999)	Qualitative	MW, Structural fragments/group contribution	Partial Least Square Discriminant Analysis
Yang *et al.(Yang et al., 2006a)*	Qualitative	Energy of highest occupied molecular orbital, total energy, molar refractivity, Log K_{ow}, Standard Gibbs free energy	Multiple Linear Regression and Back Propagation Artificial Neural Network
Gamberger et al.(Gamberger *et al.*, 1996a; Gamberger *et al.*, 1996b)	Qualitative	Structural features, molecular weight	Inductive machine learning
Ceriani et al (Chirico and Gramatica, 2011)	Qualitative	DRAGON and PaDEL descriptors	Classification and regression tree (CART) and k-nearest neighbours (kNN)
Mansouri et al (Mansouri *et al.*, 2013)	Qualitative	DRAGON descriptors	k-nearest neighbours (kNN), Partial Least Square Discriminant Analysis and support vector machines (SVM)
Paris *et al.* (Paris *et al.*, 1983)	Quantitative	Van der Waal's radii	Simple Linear Regression
Paris and Wolfe (Paris and Wolfe, 1987)	Quantitative	Van der Waal's radii	Simple Linear Regression
Paris et al. (Paris *et al.*, 1982)	Quantitative	Van der Waal's radii	Simple Linear Regression
Banerjee *et al.(Banerjee et al., 1984)*	Quantitative	Log K_{ow}	Simple Linear Regression
Okey and Stensel (Okey and Stensel, 1996)	Quantitative	Connectivity index, Chemical groups, dummy variables	Multiple linear regression
Pitter (Pitter and Chudoba, 1990)	Quantitative	Hammett Constant	Simple Linear Regression
Lee and von Gunten (Lee and von Gunten, 2012)	Quantitative	Hammett/Taft sigma Constant	Simple Linear Regression

2.6 Biodegradation processes - a 'black box' for ecological modelling programs

The microbial removal of pollutants from the environment is influenced by multiple parameters and is poorly understood. The chemical concentration, its structure, environmental conditions and activity of degrading microorganisms are key factors that determine the extent of biodegradation in any systems (Pavan and Worth, 2008; Itrich *et al.*, 2015; Kowalczyk *et al.*, 2015b). However, the biodegradation of a chemical is viewed with respect to its intrinsic properties of chemicals, neglecting extrinsic factors, especially the activity of degrading microorganisms. The activity of the microbial population is mostly dependent on the species initially present in the environment, their relative population densities, induction of their enzymes and ability to grow in the environment after being exposed to the pollutant. Current biodegradation regulatory tests however, do not consider the microbial inocula in the test system (Kowalczyk *et al.*, 2015b). In addition, little is known about the key chemical degraders and the potential chemical biodegradation pathways, and how they are related to the complexity and dynamism of different environmental compartments.

The diversity of microorganisms and catabolic enzymes having the potential to degrade chemicals in the environment is overwhelming (Grady, 1985; Wackett, 2009). A group of bacteria present in one site (e.g. soil, surface water, sewage treatment plant, sediment) varies from other sites. Likewise, even in the same site, it can vary over time (Kowalczyk *et al.*, 2016). As a consequence, the biodegradation data from regulatory tests conducted with inocula from environmental systems are poorly reproducible. In such test systems, the physical and chemical concentration (i.e. the spiked test chemical) might exert selective ecological pressure on the microbial community of the inoculum, and as a result various species within the community react in different ways; species either starve and die, remain senescent or enrich and grow.

2.6.1 Microbial ecology in biodegradation testing

In recent years, advanced high throughput next – generation sequencing methods (i.e. single gene amplicon sequencing, metagenomics and transcriptomics) and other culture independent molecular techniques have been used extensively to analyse environmental samples (Moran, 2009; de Menezes *et al.*, 2012; Jung *et al.*, 2016; Kowalczyk *et al.*, 2016). These molecular techniques provide a direct means for the in-depth characterization of diversity and function of microbial communities in environmental systems. Molecular tools are becoming affordable and are mostly based on direct assessment of nucleic acids extracted from environmental samples, and thus can be used alongside traditional biodegradation test methods to study chemical biodegradation.

Metagenomics is a study of DNA recovered directly from an environmental sample and allows sequencing of DNA as a single unit in a culture independent manner, and provides an opportunity to characterise the microbial diversity of a sample (Kowalczyk *et al.*, 2015b; Jung *et al.*, 2016). Transcriptomics allows the study of functional and metabolic diversity of the microbial community in a sample (Moran, 2009; de Menezes *et al.*, 2012; Kowalczyk *et al.*, 2015b). In general, these approaches can be applied as tools to identify and quantify the key putative chemical degraders and their activities by monitoring the major catabolic genes (e.g., monooxygenases and dioxygenases genes) involved in the biodegradation pathways of chemical degradation. Thus, the presence of key catabolic genes, encoding enzymes with the potential to degrade such chemicals, indicates the potential of a microbial inoculum for chemical biodegradation. This can be a potential metric to evaluate the biodegradability of a chemical and its fate in environment.

2.6.2 Role of catabolic genes in aerobic biodegradation of chemicals

Biodegradation of aromatic chemicals under aerobic conditions has received more attention, as it is more thermodynamically favourable than biodegradation under anaerobic conditions (Meynet *et al.*, 2015). The aerobic metabolism of aromatic chemicals commonly initiates with the incorporation of molecular oxygen into the aromatic ring by monooxygenase or dioxygenase enzymes, forming dihydrodiol (catechol moieties) with subsequent aromatic ring cleavage (Pitter and Chudoba, 1990). These reaction steps are often considered the rate-limiting step in biodegradation pathways, with genes involved in these reaction considered as key catabolic genes (Pitter and Chudoba, 1990).

The key catabolic genes have been used as molecular markers to reveal the diversity of microbial populations degrading such chemicals. Various studies have been conducted to study the role of such genes during the biodegradation of aromatic chemicals in different environmental compartments (Zhang *et al.*, 2008; Lillis *et al.*, 2010; Kowalczyk *et al.*, 2015a). Changes of key catabolic genes, their expression (catechol 2,3-dioxygenase, chlorocatechol 1,2-dioxygenase [tfdC]) and changes in microbial community structure were studied by Lillis et al. (Lillis et al., 2010) during the degradation of 2,4-dichlorophenol in soil microcosms. Likewise, Zhang et al. (Zhang et al., 2008) studied, how the key catabolic genes (i.e. catechol 2,3-dioxygenase, catechol 1,2-dioxygenase and alkane-catabolic genes [alk]) and the microbial community structure of inocula changed over the duration of the biodegradation of nonylphenol ethoxylates (NPEOs) and nonylphenol (NP) in assays using natural water microcosms. Similarly, Kowalczyk et al. (Kowalczyk *et al.*, 2015a) used *pnpA* and *mar* genes, catabolic genes involved in para-nitrophenol (PNP) biodegradation pathways to study the

diversity of the PNP degrading bacterial population in river water. These studies support the application of functional genes in monitoring the potential of any environmental inoculum for chemical biodegradation and thus can be considered as a supporting tool in biodegradation tests. Furthermore, the change or shift in microbial community composition due to the adaptation of putative microbial degraders can be identified. This can be achieved by correlating the abundance and diversity of specific chemical degraders with the fate of the chemical in that environmental compartment.

2.7 Research Gaps

Most of the currently available quantitative models do not comply with the OECD guidelines (see above; Section **2.5.3**); they are built based on a small set of congeners with a undefined, or at least uncertain, applicability domain (Paris *et al.*, 1982; Paris *et al.*, 1983; Banerjee *et al.*, 1984; Paris and Wolfe, 1987; Pitter and Chudoba, 1990; Okey and Stensel, 1996) and rarely provide potential mechanistic information on the biodegradation process. This limits their application for predicting the biodegradability of chemicals with diverse structural properties. In addition, QSBRs developed using biodegradation data from RBT results fundamentally undermines current efforts to reliably predict half-lives, as RBTs are just pass/fail tests, and do not provide accurate information on rates, and certainly not on half-life endpoints used for persistence assessments. Hence, it is imperative to validate such models against experimentally determined biodegradation rates using known degraders in simplified constrained systems, at least with rates for several representative chemicals among the set of chemicals used for QSBR model development.

Biodegradation is viewed with respect to the intrinsic properties of chemicals, while measurement of the presence and activity of potential degraders in microbial communities receives little practical attention when investigating the specific chemical degradation by mixed microbial communities. Therefore, there is a need of techniques that should be used in parallel with biodegradation test to determine not only the overall size of the inoculum, but also its quality (number of specific degraders) and the diversity of functional communities involved in the biodegradation process. 16S rRNA amplicon and functional gene analyses of inocula used in the biodegradation test will help to evaluate the relationship between microbial diversity and biodegradation outcomes, the occurrence and abundance of specific degraders and the shift in the bacterial population as a whole. This will ultimately enable better understanding of the fate of chemicals as a function of the microbial communities present in an environmental compartment and help in better prediction of chemical biodegradation.

2.8 Aims and Objectives

Aims

1. To develop a rational framework to establish and calibrate QSBR models in accordance with OECD principles in order to predict biodegradation rates and ultimately half-life of a given chemical, which is the endpoint used for chemical persistent assessment
2. To explore how molecular microbial tools can be used in parallel with biodegradation test, in order to better interpret biodegradation data, which can contribute in the prediction of chemical fate in the environment.

In order to achieve the aforementioned aims, following objectives were set:

a. To develop and validate QSBR models with a diverse set of aromatic chemicals according to OECD principles for QSBR validation.
b. To verify and calibrate the developed QSBR model with experimentally determined biodegradation rates using known degraders in simplified constrained systems, for several representative (test) chemicals among the set of chemicals used for QSBR model development
c. To design, validate and optimize molecular tools to assess the genetic potential of environmental inocula for test chemical biodegradation.
d. To develop and validate a statistical model to predict the biodegradation rate of test chemicals with initial chemical concentration and putative degrader numbers in the inocula.
e. To combine traditional biodegradation test with 16S amplicon sequencing and gene analysis techniques to study the biodegradation of chemicals, where the influence of chemical structure, its concentration and inocula concentration (i.e. number of specific degraders) on inocula microbial diversity and biodegradation rates are evaluated.

Chapter 3

A statistical modelling approach to predict biodegradation rates of aromatic chemicals.

Chapter 3. A statistical modelling approach to predict biodegradation rates of aromatic chemicals.

3.1 Abstract

The objective of this work was to develop a QSBR (Quantitative Structure Biodegradation Relationship) model for the prioritization of organic pollutants based on biodegradation rates from internationally harmonized biodegradation tests using relevant molecular descriptors that provide potential mechanistic insights into biodegradation. To do this chemicals were first categorized into three groups (Group 1: simple aromatic chemicals with a single ring, Group 2: aromatic chemicals with multiple rings and Group3: aromatic chemicals from both Group 1 and Group 2) based on molecular descriptors (those included in the DRAGON software), first order biodegradation rate of the chemicals were estimated using ultimate biodegradation rating values derived from the BIOWIN3 model, and finally developed, validated and defined the applicability domain of models for each group using a multiple linear regression approach. All the developed QSBR models complied with OECD principles for QSAR validation. The biodegradation rate in the models for the two groups (Group 2 and 3 chemicals) are associated with abstract molecular descriptors of molecular geometry, stereochemistry, conformational index and 2D finger printing (Group 2; R^2 = 0.8647, Q^2_{loo} = 0.8176 and Group 3; R^2 = 0.8156, Q^2_{loo} = 0.7905) that provide little relevant information on the mechanisms of different rate-limiting steps associated with the biodegradation process. However, the molecular descriptors associated with the QSBR model for Group 1 chemicals (R^2 = 0.8924, Q^2_{loo} = 0.8718) provided information on properties that can be linked to the underlying biodegradation process. In combination, these results lead to the conclusion that QSBRs can be an alternative tool to estimate the persistence of chemicals, some of which can provide further insights into the principles underlying biodegradation.

3.2 Introduction

Quantitative Structure Activity Relationships (QSARs) are a widely used modelling technique in diverse fields where the physicochemical properties of chemicals (independent variable or molecular descriptor) are correlated to the strength of a given response or activity (dependent variables)(Okey and Stensel, 1996; Roy *et al.,* 2011; Lee and von Gunten, 2012; Mikolajczyk *et al.,* 2015). In recent years, the application of QSARs to biodegradation (Quantitative Structure Biodegradability Relationship, QSBR) has been advocated (Pavan and Worth, 2008; Rücker and Kümmerer, 2012), providing valuable prediction of relative or absolute biodegradation rates, and/or transformation products of biodegradable chemicals, strictly on the basis of chemical structures without having to undertake laboratory testing. Hence, such correlations could provide a powerful tool to forecast the environmental fate of chemicals and

assist in ranking them for in-depth evaluation. However, such techniques are either rarely used in practice or rarely receive regulatory acceptance (with QSARs providing reliable regulatory data for 20 of 15,873 registered chemicals (ECHA, 2017) (Accessed Date:23/06/2017), largely due to a lack of high-quality experimental data (Rücker and Kümmerer, 2012).

Biodegradation models are broadly classified into qualitative and quantitative models. Qualitative models simply predict if a chemical is biodegradable or not, mostly using structural fragments as a molecular descriptor (Gamberger *et al.*, 1996a; Gamberger *et al.*, 1996b; Loonen *et al.*, 1999; Tunkel *et al.*, 2000). Whereas, quantitative models predict the extent and sometimes biodegradation pathways of chemicals, mostly using different structural and physicochemical properties of a chemical (such as molecular descriptors). The rigorous validation of quantitative models such as QSBRs requires both a defined endpoint and applicability domain, should be associated with appropriate measures of goodness of fit, robustness and predictability, and should be expressed in the form of an unambiguous algorithm capable of providing a potential mechanistic interpretation of biodegradation if possible, as per Organization for Economic Co-operation & Development (OECD) principles for QSAR validation (Netzeva *et al.*, 2005). Most of the currently available quantitative models do not comply with the guidelines; they are built based on a small set of congeners without a defined, or at least an uncertain, applicability domain (Paris *et al.*, 1982; Paris *et al.*, 1983; Banerjee *et al.*, 1984; Paris and Wolfe, 1987; Pitter and Chudoba, 1990; Okey and Stensel, 1996), and rarely provide potential mechanistic information on the biodegradation process. This limits their application in predicting the biodegradability of chemicals with diverse structural properties. There is a call for the development of new validated and mechanistic quantitative models (Pavan and Worth, 2008). This will depend on the increasing availability of high quality endpoint data for biodegradation (preferably quantitative) for a diverse set of chemicals.

Advances in computational and statistical tools have allowed researchers to move towards the development of more sophisticated quantitative biodegradation models by allowing the calculation of the numerous 1D, 2D, 3D – structural and quantum mechanical molecular descriptors that can provide potential information on a chemical's structural, physical, or electronic properties that have influence on biodegradation process (Mauri *et al.*, 2006; Helguera *et al.*, 2008; Pavan and Worth, 2008). The availability of databases containing internationally harmonized biodegradation data further allows such models to be tested and validated (Mansouri *et al.*, 2013; Ceriani *et al.*, 2015). In this chapter, QSBR models for three groups of aromatic chemicals (simple [**Table A 2, Appendix A**], complex [**Table A 3,**

Appendix A], and a combination of simple and complex aromatic chemicals [**Table A 4, Appendix A**]) were developed and validated (according to OECD principles for QSAR validation). A strong emphasis was given on those molecular descriptors that could be easily measured and/or interpreted in terms of a potential mechanism for biodegradation. For this purpose, we initially selected 140 (**Table A 5, Appendix A**) aliphatic, aromatic and cyclic chemicals of environmental concern.

3.3 Methods

3.3.1 Chemical Selection and Molecular Descriptors Choice

140 organic chemicals (**Table A 5, Appendix A**) categorized either as priority pollutants or emerging organic pollutants in the field of water policy were initially selected for model development (No, 2001; EPA, 2003; Lee and von Gunten, 2012). The identified chemicals encompass aliphatic, aromatic and cyclic chemicals.

Each chemical was characterized by 4897 molecular descriptors (12 descriptors from online databases and 4885 computed by Dragon software [(Dragon); Accessed Date: 22/03/2015]). The descriptors of selected chemicals in the Dragon software (Dragon) were computed with the optimized structure (i.e. structure with minimum energy conformation) of the chemicals. Descriptors with constant values (i.e. some descriptors have the same value for all chemicals) were excluded in a pre-reduction step; thus obtaining a set of 2459 Dragon molecular descriptors. Detailed information on Dragon molecular descriptors can be found in the Handbook of Molecular Descriptors (Todeschini and Consonni, 2008).

3.3.2 Endpoint for QSBR model

Endpoint refers to any physicochemical, biological or environmental effect that can be measured and therefore modelled. Most of the biodegradation data available in the literature are based on data-poor experimental information and are rarely reproducible (Rücker and Kümmerer, 2012). This deficit demands for the reliable biodegradation data obtained from the globally accepted standard biodegradation test. In addition, the development of a mechanistic QSBR model requires a defined endpoint that is consistent, reliable and globally accepted. Halve-lives which are directly linked to first-order rate, are the end-point commonly used in the regulatory assessment of persistence (Pavan and Worth, 2008). First order biodegradation rates derived from BIOWIN3 ultimate biodegradation rating were used as a model endpoint. BIOWIN3 is one of several environmental fate estimation models incorporated in the EPI (Environmental Protection Interface) Suite and uses internationally harmonized biodegradation data (EPA, 2011). It predicts relative ultimate biodegradation rates of chemicals using the fragment based additive approach (EPA, 2011).

In this study, a new statistically significant regression relating the ultimate biodegradation rating and ultimate biodegradation half-life of chemicals (eChemPortal) was developed [**Table A 1(A and B) and Figure A 1, Appendix A**], using selected chemicals incorporated in the BIOWIN3 model;- this regression was used to convert semi-quantitative BIOWIN3 biodegradation ratings to half-lives for the selected chemicals in the study. The chemicals having the highest, lowest, and values between the highest and lowest values of ultimate biodegradation rating, were used to develop the regression. Finally, the corresponding first order biodegradation rate of chemicals used in this work was computed using the degradation half-life of the chemicals [**Equation 1, Chapter 2**]. The natural logarithm of first order rate was used as the response variable for subsequent QSBR modelling.

3.3.3 Screening of Chemicals for QSBR model development

Clustering analysis enables pattern recognition and natural grouping of the samples into clusters, which are not known beforehand, by using the common property characterized by the values of a set of variable (Pirhadi *et al.*, 2015). In this study, Hierarchical Clustering Analysis (HCA) was used to group chemicals into clusters based on calculated molecular descriptors (Pirhadi *et al.*, 2015). This resulted four main clusters, composed, respectively, of aliphatic chemicals and few simple aromatic chemicals (cluster1); simple chemicals having a single aromatic ring and a few aliphatic and acyclic chemicals (cluster2 and 3); and aromatic chemicals having more than one aromatic ring and a few cyclic chemicals (cluster4) [**Figure A 3, Appendix A**]. Aliphatic chemicals from the first cluster were excluded from further analysis as there were too few (20 chemicals) for reliable QSBR model development and validation. Furthermore, few acyclic and aliphatic chemicals were present in clusters 2, 3 and 4, which were therefore also excluded from further analyses. Therefore, only aromatic chemicals were considered for further model development.

3.3.4 Dataset Splitting

The selected aromatic chemicals were classified into three groups (see **Table A 2, A 3 and A 4, Appendix A**) [Group1 (Simple aromatic chemicals with one aromatic ring); cluster 1+cluster2 +cluster3, Group2 (Aromatic chemicals with multiple aromatic rings): cluster 4 and Group3 (Group 1 + Group 2); cluster1+cluster2+cluster3+cluster4]; each group was analysed as a separate dataset for QSBR modelling. Firstly, Principal Component Analysis [PCA] was used for each group to identify any outliers that could affect the robustness and fitness of the model (Gramatica *et al.*, 2013) (Pirhadi *et al.*, 2015). This analysis also aided in selecting training and validation sets. Subsequently, the random-by-response approach (sorting chemicals by ordering them according to increasing/decreasing order of end point

value) was applied to split the chemicals into training and validation sets (Mikolajczyk *et al.*, 2015). To perform splitting, the third chemical was selected from the set as a first validation chemical. Subsequently every third chemical (if possible) was selected from the sorted list as validation chemical. The remaining chemicals formed the training set. The chemical with the highest and the lowest biodegradation rates were included in the training set, to guarantee that the prediction set spanned the entire range of the experimental measurements and was numerically representative of the dataset. The final QSBR model for Group1, Group2 and Group3 were developed with 60, 28 and 84 chemicals respectively.

3.3.5 QSBR Model Development and Validation

QSARINS software (Gramatica *et al.*, 2013) was used to develop a QSBR model for the split dataset using Multiple Linear Regression (MLR) techniques. In MLR, the endpoint (y_i) is described with the best combination of the most relevant auto-scaled descriptors used as independent variables (x_1, x_2,…x_n), as follows:

$$y_i = b_0 + b_1x_1 + b_2x_2 + \cdots . + b_nx_n \quad (1)$$

where b_0 is the intercept and b_1, b_2,… b_n are the regression coefficients.

The best combination of the most relevant descriptors was selected using the Genetic Algorithm (GA) (Pavan and Worth, 2008; Gramatica *et al.*, 2013) incorporated in QSRINS. This technique allows identification of the best solution (i.e. helps to search the best combination of descriptors) by maximizing (or minimizing) a selected fitness function. In this study, Q^2_{Loo} (Cross validated coefficient; used to evaluate the model's performance in predictions) was selected as a fitness function. **Figure 3-1** summarizes the detailed methodological steps performed during QSBR model development.

The best models developed by QSARINS were sorted using fitting (R^2, Appendix A, Eq.1 and $RMSE_{tr}$, Appendix A, Eq.2) and robustness (Q^2_{LOO}, Appendix A, Eq.3 and $RMSE_{CV}$, Appendix A, Eq.4) criteria. Correlation coefficient (R^2) and the root mean square error of calibration (RSMEC) were used as measures of the goodness of fit for the developed model (Puzyn *et al.*, 2009). While, cross validated coefficient Q^2_{LOO} (leave one out method) and root-mean-square-error of cross validation (RMSECV) were used to verify its stability and robustness (Puzyn *et al.*, 2009). For details, please refer to **Appendix A**.

The internally optimized, stable and robust models were further evaluated for their external predictive power with chemicals not used in the model building process using different external validation parameters like Q^2_{F1} (Eq.5, Appendix A), Q^2_{F2} (Eq.6, Appendix A), Q^2_{F3} (Eq.7, Appendix A) and root-mean-square-error of prediction ($RMSE_P$, Eq.8) (Chirico and Gramatica, 2011). These are predictive squared correlation coefficient (for details please refer to Appendix A).

In addition, the Applicability Domain (AD) [the theoretical area of the chemical space, where for the particular mechanism of biological action or function, the model's predictions are reliable](Puzyn *et al.*, 2009) of the finally selected model was assessed by the leverage approach and using the Williams graph (Roy *et al.*, 2011), a plot of leverage values (h) versus standardized residuals that generally identifies the structural outliers (X-Outliers, those having leverage value greater than critical h value) and the residual outliers (Y-outliers, those with predicted response value above the user defined standardized residual limit). The critical h value (h^*) is calculated as:

$h^* = 3(p+1)/n$, where p is the number of model predictors, and n is the number of objects (training chemicals) used to calculate the model.

The leverage value (h_i) is calculated from the molecular descriptors included in the model, and estimated according to Equation 9 (Appendix A).

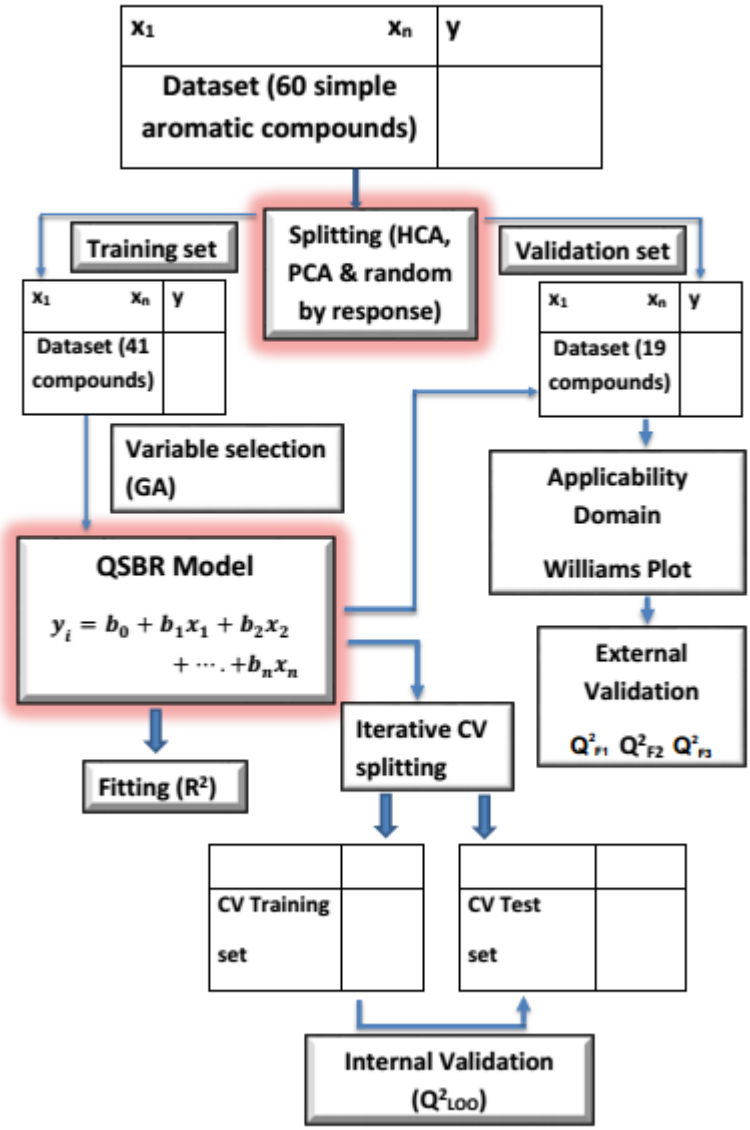


Figure 3-1 Detailed procedure of developing QSBR (adopted from (Mikolajczyk *et al.*, 2015))

After identifying the best model with three Dragon descriptors for Group1 chemicals, an attempt was made to develop new models that incorporated quantum mechanical descriptors and descriptors describing different physicochemical and structural properties of chemicals (**Table A 6, Appendix A**). These descriptors were those that pertained to electronic properties and functional groups within the chemical, which have been shown to influence the biodegradation of a chemical (Pitter and Chudoba, 1990; Nolte and Ragas, 2017). The quantum mechanical descriptors were calculated at the semi-empirical level of theory with the use of PM7 method implemented in MOPAC2012 software (Stewart, 1994), whereas other descriptors were obtained from different online databases (Ballabio *et al.*, 2009). A similar approach as described above was used for model development.

3.4 Result and Discussion

3.4.1 QSBR models

Table 3-1 provides the overall summary of the best model for each set of chemical groups. The QSBR model for the simple aromatic chemicals (**Equation 4**) was better than the other two models for complex (**Equation 3**) and aromatic chemicals (**Equation 2**), as indicated by higher values for R^2, Q^2_{loo} $Q^2_{F1,}$ Q^2_{F2} and Q^2_{F3}. Equation 2, 3 and 4 are the model equation for the three best models for each group and description of each model descriptor are provided in **Table 3-1.**

$$\mathbf{Ln(rate)} = \mathbf{-2.6767 - 0.0408 * Sp\_Abs\_B(m) - 0.9678 * P2u + 0.0039 * P\_VSA\_e\_5} \quad (2)$$

$$\mathbf{Ln(rate)} = \mathbf{3.226 - 1.6447 * SM2\_B(m) + 0.0803 * RDF145e} \quad (3)$$

$$\mathbf{Ln(rate)} = \mathbf{-3.2495 - 0.2237 * nN - 0.3907 * Mor08u - 0.2781 * nArX} \quad (4)$$

The molecular descriptors provide information on specific physicochemical or structural characteristics of the chemical, and the ability to interpret the encoded value of the descriptors provide information on the molecular features that are most likely to effect the biological activity of the studied chemical (Todeschini and Consonni, 2008). However, when using an extensive matrix of molecular descriptors, the mechanistic interpretation of the endpoint of interest may not always provide useful or easily interpretable information. This is evident in the current study, as the descriptors associated with QSBR models for chemical datasets belonging to Group 2 and Group 3 are relatively abstract molecular descriptors of molecular geometry, stereochemistry, conformational index, 2D finger printing and fragments counts. While for Group 1, the descriptors are for hydrophobic, electronic, steric, size and shape properties of chemical and can be more easily interpreted with respect to potential mechanistic explanations of their effect on biodegradation. Therefore, in upcoming sections, the QSBR model for simple aromatic chemicals was principally focused on; evaluating the model in terms of different model parameters, defining its applicability domain and providing the mechanistic linkage between the biodegradation rate and the molecular descriptors associated with the model. Further, an attempt was made to improve the existing model by incorporating some other common descriptors specifically associated with biodegradation of chemicals, namely, quantum mechanical, hydrophobic, steric and electronic descriptors (**Table A 6, Appendix A**)

Table 3-1 Summary of the Best QSBR model for each of the selected groups of chemicals

Chemical Group	Total Number of Chemicals in the model	Model Equation	Model Statistics	Description of each descriptor used in the model
Aromatic chemicals (Group3)	**Total: 84** Training Set: 59 Validation Set: 25	$\mathbf{Ln(rate)} = -2.6767 - 0.0408 * \mathbf{Sp\_Abs\_B(m)} - 0.9678 * \mathbf{P2u} + 0.0039 * \mathbf{P\_VSA\_e\_5}$	$\mathbf{R^2} = 0.8156$ $\mathbf{Q^2_{loo}} = 0.7905$ $\mathbf{Q^2_{F2}} = 0.8138$ $\mathbf{Q^2_{F1}} = 0.8130$ $\mathbf{Q^2_{F3}} = 0.8157$	**Sp_Abs_B(m):** graph energy from Burden matrix weighted by mass (2D matrix-based descriptors) **P2u** : 2nd component shape directional WHIM index / unweighted (WHIM descriptors)[a] **P_VSA_e_5** : P_VSA-like on Sanderson electronegativity, bin 5 (P_VSA-like descriptors)[b]
Complex Aromatic Chemicals (Group2)	**Total : 28** Training Set: 18 Validation Set: 10	$\mathbf{Ln(rate)} = 3.226 - 1.6447 * \mathbf{SM2\_B(m)} + 0.0803 * \mathbf{RDF145e}$	$\mathbf{R^2} = 0.8647$ $\mathbf{Q^2_{loo}} = 0.8176$ $\mathbf{Q^2_{F2}} = 0.8541$ $\mathbf{Q^2_{F1}} = 0.8373$ $\mathbf{Q^2_{F3}} = 0.8859$	**SM2_B(m)**: spectral moment of order 2 from Burden matrix weighted by mass (2D matrix-based descriptors) **RDF145e** : Radial Distribution Function - 145 / weighted by Sanderson electronegativity (RDF descriptors)[c]
Simple Aromatic Chemicals (Group1)	**Total : 60** Training Set : 41 Validation Set : 19	$\mathbf{Ln(rate)} = -3.2495 - 0.2237 * \mathbf{nN} - 0.3907 * \mathbf{Mor08u} - 0.2781 * \mathbf{nArX}$	$\mathbf{R^2} = 0.8924$ $\mathbf{Q^2_{loo}} = 0.8718$ $\mathbf{Q^2_{F2}} = 0.8829$ $\mathbf{Q^2_{F1}} = 0.8835$ $\mathbf{Q^2_{F3}} = 0.9178$	**nN** : Number of nitrogen atom (Constitutional Index) **Mor08u** : un-weighted descriptor with scattering parameter (s) = 7 $Å^{-1}$ (3D Molecular Representations of Structure based on Electron diffraction) **nArX** : Number of halogen on aromatic rings (Functional group counts)

[a] WHIM descriptors; Weighted Holistic Invariant Molecular descriptors) are geometrical descriptors based on statistical indices calculated on the projections of the atoms along principal axes

[b] P_VSA-like descriptor; the amount of van der Waals surface area (VSA) having a property P in a certain range

[c] RDF descriptor; Molecular descriptors obtained by radial basis functions centered on different interatomic distances (from 0.5A to 15.5A)

3.4.2 QSBR model for simple aromatic chemicals (Group 1)

Table 3-3 provides the summary statistics for the three-descriptor based QSBR model for simple aromatic chemicals (Equation 12). This model shows high stability ($R^2 = 0.8924$), robustness ($Q^2_{LOO} = 0.8718$), and external predictive ability ($Q^2_{F1} = 0.8829$, $Q^2_{F2} = 0.8835$, and $Q^2_{F3} = 0.9178$). The plots of the experimental versus predicted values (**Figure 3-2**, A and B) showed very good agreement between BIOWIN3 derived first order biodegradation rates and the model predicted values of biodegradation rate for 60 aromatic chemicals for both training and validation sets. Likewise, model predicted half-life and BIOWIN3 derived half-life also showed good agreement. In addition, the plots confirmed the predictive capability of the developed model.

The Williams plot (**Figure 3-3**) verified the absence of outliers (residual values (y_i -$\hat{y}_i$) were within the limits ± 3 times standard deviation), and showed good applicability of the model for the prediction of the biodegradation rates for all of the studied mono-aromatic chemicals. In addition, none of the structures of the studied aromatic chemicals were substantially different from the training set chemicals; showing a leverage value $h_i < h^* = 0.30$, proving the applicability of this model for untested mono-aromatic chemicals with calculated h_i values lower than the critical value ($h^* = 0.30$). However, it has to be noted that h_i values were estimated using **Equation 9 (Appendix A),** which does not take into account structural variability or differences other than those expressed by the selected descriptors. Therefore, the applicability of the model in the actual form is restricted to chemicals which have structural similarities to those used for the training set (i.e. mono-aromatic chemicals with different substituent patterns), and should not be employed in prediction for polyaromatic chemicals.

After identifying the best three Dragon descriptors (nN, nArX and Mor08u), a further 12 descriptors were included in the dataset to develop new models using the same approach (**Table A 6, Appendix A**). The summary of the best 10 models based on 4 descriptors is reported in **Table 3-2**. All the models are robust ($89.8 < R^2 > 93.1$ %) and stable (86.9 % > $\boldsymbol{Q^2_{loo}}$ < 91.1 %) and have good predictive ability (85.6 %< Q^2ext > 95.8 %). In addition, all the descriptors from the developed models provide mechanistic insight into the model, as these descriptors are related to electronic, steric and lipophilic properties of chemicals. Biodegradation of a chemical has been shown to be influenced by electronic and lipophilic properties, along with functional groups within the chemical (Parsons and Govers, 1990a).

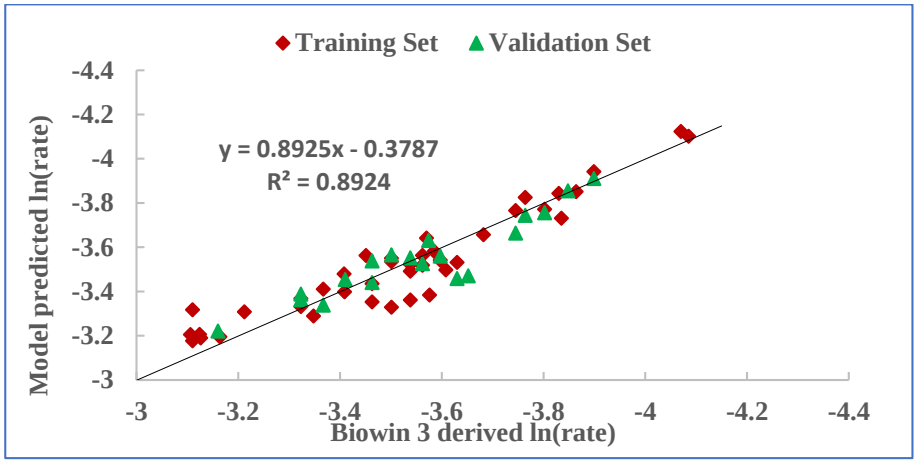
Table 3-2 Summary of the 10 best models with 4 descriptors. Models were developed using Multiple Linear Regression (MLR) technique and the best combination of the most relevant descriptors was selected using the Genetic Algorithm (GA). Molecular descriptors used in the model are either electronic, steric or lipophilic properties of chemicals.

No.	Size	Variables	R^2	$RMSE_{tr}$	F	Q^2_{loo}	RMSEcv	$RMSE_{ext}$	Q^2_{F1}	Q^2_{F2}	Q^2_{F3}
1	4	nN Mor08u nArX MR	0.931	0.067	117.203	0.911	0.075	0.053	0.937	0.937	0.956
2	4	nN Mor08u nArX vX1	0.930	0.067	115.552	0.911	0.075	0.056	0.931	0.931	0.952
3	4	nN Mor08u nArX vdw Volume	0.928	0.068	113.133	0.907	0.077	0.052	0.940	0.940	0.958
4	4	nN Mor08u nArX vX2	0.924	0.069	106.728	0.905	0.078	0.060	0.920	0.920	0.944
5	4	nN Mor08u nArX electronic energy	0.915	0.074	94.201	0.890	0.084	0.068	0.898	0.898	0.928
6	4	nN Mor08u nArX MW	0.914	0.074	93.498	0.886	0.085	0.070	0.891	0.891	0.923
7	4	nN Mor08u nArX total energy	0.909	0.076	87.780	0.883	0.086	0.073	0.881	0.880	0.916
8	4	nN Mor08u nArX σ	0.905	0.078	82.965	0.879	0.088	0.075	0.876	0.875	0.912
9	4	nN Mor08u nArX Log P	0.904	0.078	82.003	0.875	0.089	0.080	0.857	0.856	0.899
10	4	nN nArX total energy vdw Volume	0.898	0.080	77.379	0.869	0.091	0.062	0.915	0.915	0.940

Table 3-3 Model statistics of 3 descriptors QSBR model.

Variable	Coeff.	Std. coeff.	p-value	Fitting	Internal validation	External Validation		
				R^2 ($RMSE_{tr}$)	Q^2_{LOO} ($RMSE_{CV}$)	Q^2_{F1} ($RMSE_{ext}$)	Q^2_{F2}	Q^2_{F3}
Intercept	-3.2495			0.894 (0.0827)	0.8718 (0.0903)	0.8835 (0.0723)	0.8829	0.9178
nN	-0.2237	-0.4604	0					
Mor08u	-0.3907	-0.3796	0					
nArX	-0.2781	-0.9225	0					

[A]



[B]

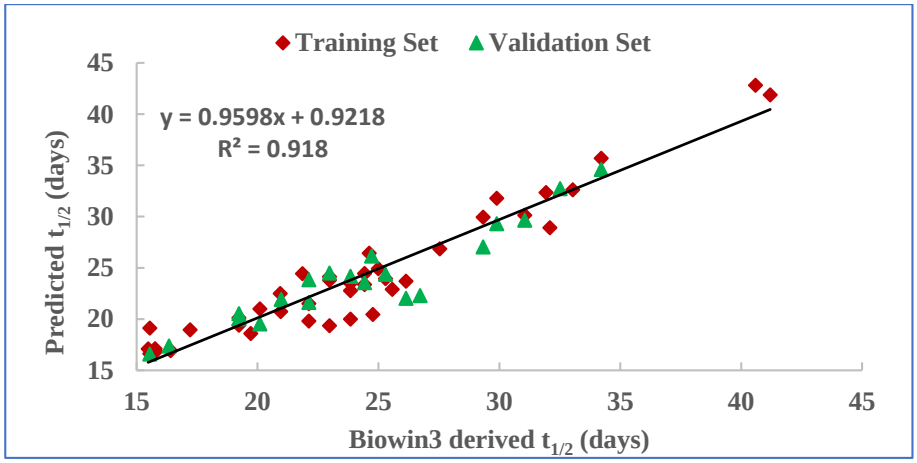


Figure 3-2 (A) Plot of BIOWIN3 derived rates versus model calculated values for the full model based on three descriptors. (B) Plot of BIOWIN3 derived half-life versus model estimated half-life for the full model based on three descriptors.

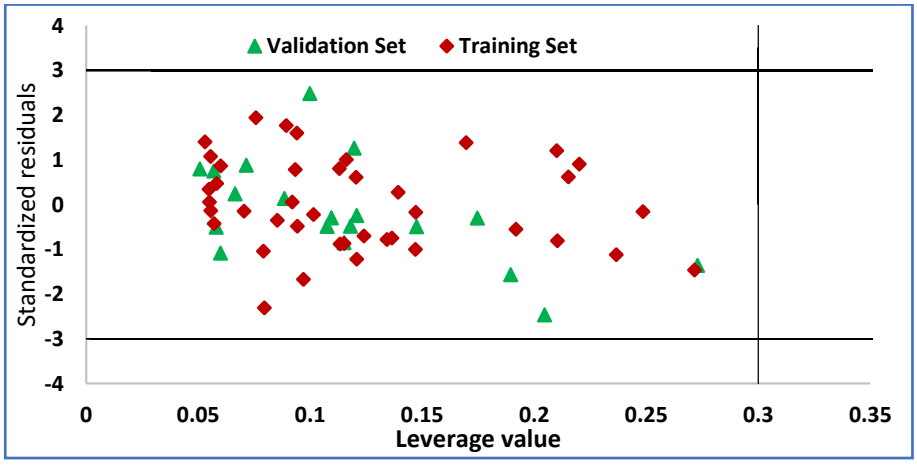


Figure 3-3 Williams plot for the applicability domain of the full model based on three descriptors

3.4.3 QSBR model descriptors and their mechanistic interpretation

The type, number and position of substitutions on the aromatic ring are important parameters that determine the electronic characteristics of aromatic chemicals and ultimately influence

the degree and rate of biodegradation. The information provided by the descriptors (substituent constant [σ], nN, nArX and Mor08u) used in this study is in agreement with the effect of substitutions in the aromatic ring as follows.

Mor08u is an un-weighted MoRSE (Molecular Representations of Structure based on Electron diffraction) descriptors with scattering parameter (s) = 7 $Å^{-1}$ and calculated according to Equation 10 (Appendix A). The calculated values of different MoRSE descriptors for chemicals are based on their structural features, where distance among the different atoms within a molecule will be the principal means to separate the molecule from others (Devinyak *et al.*, 2014). When QSAR models are developed for a structurally similar group of chemicals, the difference in the value of 3D-MoRSE descriptor is due to several neighboring atom pairs in the molecule (Devinyak *et al.*, 2014). The differences in these descriptor values result in different physicochemical properties for the monoaromatic chemicals, attributable to differences in the number, type and position of substituents on the aromatic ring, which in turns affects the biodegradation rate. This interpretation is in agreement with results previously published in the literature (Sikarwar and Dixit, 2012), where an individual QSAR model was used to predict the molar refractivity, partition coefficient and polarizability of 34 phenolic chemicals, using 3D-MoRSE descriptors and Eigen values respectively.

Number of Nitrogen (nN) and Halogen Atoms in Aromatic Chemical (nArX): The biodegradability of a molecule is influenced by electronic (inductive and mesomeric), steric and lipophilic properties of a molecular system (Pitter and Chudoba, 1990). In particular, the type, number and position of substituent groups present in the aromatic molecule strongly determine the electronic characteristics of that molecule. Both electronic inductive and mesomeric effects are responsible for attraction or repulsion of electrons, affecting the electron density in the reaction center. The presence of halogen and nitrogen atoms in the aromatic system will contribute to deplete the electron density of the reaction system. In biodegradation reactions, the initial attack on the aromatic ring is mostly assumed to be electrophilic in nature, therefore the presence of electron – attracting elements (like halogens and nitrogen) deactivate the aromatic ring in certain positions for attack by oxygenases and results in lower a biodegradation rate (Pitter and Chudoba, 1990).

3.4.4 Mechanism of Aerobic Biodegradation

The most practical and useful approach to QSBRs is to relate biodegradation rate data to molecular descriptors specifically providing relevant information on mechanisms of the biodegradation rate-limiting process. Furthermore, biodegradation rates are believed to be a

function of the rate of a series of processes that occurs in a stepwise manner (Parsons and Govers, 1990a).

Extracellular enzymes initiate microbial mineralization of organic matter by hydrolyzing substrates to sizes sufficiently small to be transported across cell membranes. They then diffuse through the cytoplasm to reach the enzyme where the biodegradation reaction is initiated (Arnosti, 2011). Subsequently, enzyme induction occurs, which ultimately results in the binding of the metabolic enzyme and the chemical followed by transformation of the chemical by that enzyme(Parsons and Govers, 1990a; Wammer and Peters, 2005) .

The transport of chemicals across the microbial membrane can occur either via active transport mechanisms or passive diffusion. An active transport mechanism is mostly involved in efflux and has little significant influence on biodegradation rate (Parsons and Govers, 1990a). It has been shown that the transport of polyaromatic hydrocarbons (PAHs) across the microbial membrane generally occurs via the passive diffusion processes in some PAH-degrading bacteria (Bugg *et al.*, 2000). Furthermore, uptake via passive transport tends to correlate with the descriptors that normally describe the hydrophobicity of the chemicals such as LogP (octanol-water partition coefficient), polarizability or molar refraction (Banerjee *et al.*, 1984; Parsons and Govers, 1990a); suggesting diffusion of a chemical through the cytoplasm is an important step in biodegradation.

The structure of chemicals also likely controls the process of metabolic enzyme induction, subsequent binding of the chemical to enzymes, and transformation of the chemical (Parsons and Govers, 1990a). The relationship between biodegradability of aromatic chemicals and molecular descriptors associated with the structure of chemicals are discussed in several QSBR models (Wolfe *et al.*, 1980; Okey and Stensel, 1996). More specifically, in mono-aromatics, the number, position and type of substituent are associated with electronic properties of the chemicals and influence their biodegradability (Pitter and Chudoba, 1990). In such chemicals, aerobic biotransformation is generally initiated with the addition of molecular oxygen to the aromatic ring (i.e. hydroxylation) by oxygenase or dioxygenase enzymes (Pitter and Chudoba, 1990; Peijnenburg, 1994), followed by aromatic ring cleavage. Hydroxylation of the aromatic ring with subsequent ring cleavage is followed by an electrophilic substitution, which are considered as the rate determining steps (Pitter and Chudoba, 1990). The presence of certain substituent groups increases the electron density of the aromatic ring and accelerates the biodegradation process. Nitrogen and halogenic substituents strongly deplete the electron density of the aromatic ring, decreasing the rate of biodegradation as compared to degradation of aromatic chemicals with other substituents (e.g., OH, CHO, CH_3). This effect was demonstrated by Alexander and Lustigman (1966), who

showed a retarded biodegradation of mono- and di- substituted benzene by soil microorganisms with chloro-, nitro- and sulfonate- substituents, whereas an increased rate was recorded in presence of hydroxyl and carboxyl group.

The shape and size of chemicals plays an integral role during fitting of a chemical into the active site of the enzyme (Wammer and Peters, 2005). Four descriptors in the current models; molecular weight, van der Waal volume and valence connectivity index (vX2 and vX1) are general size and shape descriptors. The association of these descriptors with biodegradation rate indicates that the steric properties of aromatic chemicals have a significant role to play in the biodegradation process. Several QSAR models relating biological activity and the molecular descriptors associated with steric properties have been developed elsewhere (Koch, 1982; Paris *et al.*, 1982; Okey and Stensel, 1996).

In most chemical reactions, an energy barrier exists, and it must be surmounted for the reaction to occur. Thus, kinetic and thermodynamic parameters are integral in explaining the observed differences in biodegradation. In addition, thermodynamic feasibility is also considered as an important metric to evaluate the potential of a biodegradation reaction (Finley *et al.*, 2009). Two descriptors, electronic energy and total energy of a molecule are quantum mechanical descriptors and generally provide information about energy associated with a chemical. The relationship between the biodegradability of aromatic chemicals and descriptors associated with the energy of molecules are discussed in several papers (Wammer and Peters, 2005; Yang *et al.*, 2006b). Total energy (ToE) of a molecule is the sum of total energy of all electrons (E_{el}) and repulsion energy between atomic nuclei ($E_{nuc-nuc}$) in a molecule (Stewart, 1994).

$$ToE = E_{el} + E_{nuc-nuc} \quad (5)$$

Electronic energy (E_{el}) is the sum of repulsion energy between electrons and the attraction energy between electron and atomic nuclei. According to molecular orbital theory, the total electronic energy is directly associated with total energies of the individual occupied molecular orbitals, which also provides the information of total bond energy in a molecule (Karelson *et al.*, 1996; Petrucci *et al.*, 1997). This suggests that the higher the total energy, the higher the total bond energy of a molecule. Thus, the molecule with higher total bond energy has strong attraction between electrons and the atomic nuclei. This implies that, high energy is required to degrade such a molecule. Therefore, when an aromatic chemical has higher total energy, they are more resistant to degradation.

3.4.5 Implications for predicting biodegradation half-lives for fate and hazard assessment

Determination of chemical half-lives in different environmental compartments are essential for the reliable risk assessment of chemicals, yet, they still remain the major source of uncertainty. Several researchers have made an attempt to convert BIOWIN output into half-lives and ultimately into rates (Gouin *et al.*, 2004; Arnot *et al.*, 2005; Aronson *et al.*, 2006), a similar task to that performed in this study, yet the reliability of the model prediction are still considered uncertain (Rücker and Kümmerer, 2012). Therefore, it is accepted that accurate prediction of half-lives or rates with a model can only be achieved through training the model with actually measured half-lives or rates (Rücker and Kümmerer, 2012). Nevertheless, a comparison was made between the first order rates predicted with the model developed by Arnot et al. (2005) and the one used in this study (**Figure A 2, Appendix A**). There was a significant correlation between the first order rates estimated with the aforementioned model and the first order rates used in this study (univariate regression analysis; $r^2 = 0.99$, p-value <0.05).

This study has shown that molecular descriptors associated with different chemical features can be potentially linked to biodegradation processes of chemicals and ultimately used for QSBR model development. All the QSBR models were developed as per Organization for Economic Co-operation & Development (OECD) principles for QSAR validation. Nevertheless, QSBR models for simple aromatic chemicals were robust, reliable and can be more easily interpreted with respect to potential mechanistic explanations of their effect on biodegradation than the other models developed in this study.

Regulatory frameworks have put greater emphasis on identification and prioritization of chemicals based on their environmental hazardous properties (i.e. persistence, bioaccumulation and toxicity (PBT)) rather than their environmental risk alone (Martin *et al.*, 2017b). Furthermore, their guidance also recommends that, persistence assessment of chemicals should be performed before bioaccumulation and toxicity assessments in order to avoid unnecessary animal tests, and the latter should be carried out only when the chemical is assigned as potentially persistent (ECHA, 2017). It has to be noted that half-lives are the commonly used end-point in the regulatory assessment of persistence (Rücker and Kümmerer, 2012; ECHA, 2017; Martin *et al.*, 2017b). The QSBR model developed here, for simple aromatic chemicals, has been shown to correctly predict and interpret respectively the rate and biodegradation mechanism of chemicals based on their previous biodegradability classification obtained from BIOWIN3 model. These models also enabled in generation of first order biodegradation rate of chemicals and thereby biodegradation half-life predictions. Therefore, QSBR models can be an alternative approach in both screening for, and making

definitive classification in, persistence assessments of chemicals, in this instance, mono-aromatic chemicals.

3.4.6 Future outlook

A reliable prediction of chemical biodegradation rate with a model provides an opportunity to prioritize chemicals that pose the greatest risk to the environment and humans. The main concern with respect to the utilization of predictive modelling is the quality of existing data. It has been widely accepted that a model is only good, if input data inevitably link the model and experimental work. Furthermore, it has to be noted that the majority of existing QSBR, including in this study, are built with the data [e.g., Ready Biodegradability Tests (RBTs) data] from different biodegradation databases, which themselves are another QSBR. However, existing regulatory screening test (i.e. RBTs) exhibit a number of limitations that are particularly pertinent to their use in persistence assessment (Kowalczyk *et al.*, 2015b). These limitations include (i) high levels of variation (inter-replicate, inter-test, inter-facility and temporal), (ii) large number of test fails and (iii) the arbitrary time restriction [the pass criterion (70% removal of Dissolved Organic Carbon [DOC] and 60% of Theoretical Oxygen Demand [ThOD] or Theoretical Carbon dioxide [$ThCO_2$] production) is reached in 10 days window within the 28 days period of test] (Kowalczyk *et al.*, 2015b). The use of RBTs results fundamentally undermines current QSBR efforts. Hence, the model needs to be further validated against experimentally determined biodegradation rates, at least by experimentally determining rates for several representative chemicals among the set of 60 chemicals. However, overall, QSBR models can, in principle, be useful alternative tools in the prioritization of persistent chemicals.

Chapter 4

A Biodegradation Test Method for Estimating Biodegradation Rates towards Evaluating QSBR Models

Chapter 4. A Biodegradation Test Method for Estimating Biodegradation Rates towards Evaluating QSBR Models

4.1 Abstract

Quantitative Structure Biodegradation Relationships (QSBRs) are a tool to predict the biodegradability of chemicals. The objective of this work was to generate reliable biodegradation rates for mono-aromatic chemicals in order to evaluate and verify a previously developed QSBRs model. A robust biodegradation test method was developed to measure biodegradation rates of chemicals in pure culture. Five representative mono-aromatic chemicals were selected that spanned a wide range of biodegradation rates. Aerobic biodegradation experiments were performed for each chemical in batch reactors seeded with known degraders. Chemical removal, degrader growth and CO_2 production were monitored over time. Experimental data were fitted using a full carbon mass balance model, and Monod kinetic parameters (Y, K_s and μ_{max}) for each chemical were determined. In addition, stoichiometric equations for aerobic mineralization of the test chemicals were developed. The estimated biodegradation rate was highest for m-cresol followed by phenol, toluene, 4-chlorophenol and 2,4-dichlorophenol. The theoretically estimated biomass yields and CO_2 were similar to those experimentally observed; 35 % (s.d ± 8 %) of the recovered substrate carbon was converted to biomass, and 65% (s.d ± 8 %) was mineralised to CO_2. Significant correlations were observed between experimentally determined biodegradation rates and the molecular descriptors characterizing the chemicals used in previous QSBR study, and also with the biodegradation rates predicted in a previous QSBR study. These correlations suggest that mechanistic QSBR models can be reliable and robust.

4.2 Introduction

In a previous chapter (**Chapter 3**), a mechanistic QSBR model for 60 simple mono-aromatic chemicals was developed in accordance with OECD principles, which related molecular descriptors to the first order rate of chemical biodegradation in the aqueous environment. These rates were calculated using semi-quantitative biodegradation data included in the BIOWIN3 model (EPA, 2011), which are based on the ratings given by experts to evaluate the relative biodegradability of chemicals and are unlikely to represent true biodegradation rates. The biodegradation data included in the BIOWIN model are derived from ready biodegradability tests (RBTs). RBTs are the first tier primary screening tests for assessing the biodegradation of chemicals for regulatory frameworks, and often the data from these tests are used to calculate rates for inputting to modelling efforts (Kowalczyk *et al.*, 2015b; Martin *et al.*, 2017c). However, RBTs exhibit a number of limitations [see Chapter 3, Section 3.4.6] and

are just pass/fail tests that provide no accurate information on rates and certainly not on half-life endpoints used for persistence assessments. The use of such flawed RBT results fundamentally undermines current efforts to reliably predict half-lives using QSBRs. Hence, in the first instance, it is imperative to validate such models against experimentally determined biodegradation rates using known degraders in simplified constrained systems. Reliable estimation of biodegradation kinetics depends on the test protocol, the experimental system used for the biodegradation assay, and the kinetic models used to fit the experimental data. Before attempting to predict biodegradation rates from the structure of chemicals in samples with complex biology (e.g., natural and engineered ecosystems), it would be informative to explore how well biodegradation rates can be predicted from their structures in more constrained systems, where biological complexity is minimised. Therefore, a good starting point is to work with simplified constrained systems with known degraders at high inoculum load to reduce biological complexity. In such system, the probable initiatory effect of chemical in microbial growth is minimized, while variation in observed biodegradation results can be reduced and ultimately reproducible rates can be obtained. In contrast, biodegradation data obtained from experiments performed with mixed inocula (e.g., activated sludge inocula) might be difficult to interpret due to complexity of inocula used. Experimental data from simplified systems can be fitted to a number of models, which enable estimation of biodegradation rates of chemicals (Simkins and Alexander, 1984; Pitter and Chudoba, 1990; McCarty, 2012).

The main objective of this study was to generate biodegradation rates from batch experiments with pure cultures of degraders to validate the principles of a previously developed QSBR model for simple aromatic chemicals. To ensure that the measured biodegradation rates from batch experiments with pure culture are reliable, and to explore potential uncertainties in their experimental determination, biomass concentrations and CO_2 were monitored in parallel with the substrate degradation, and a comprehensive carbon mass balance model was developed to rigorously interpret the experimental data from the batch system. Furthermore, the experimentally determined biomass yields were compared to theoretical predictions based on stoichiometric equations for microbial growth on each single substrate, to provide in-depth understanding of the biodegradation process and ensure universality. Finally, the validation and verification of a previously developed QSBR model (Chapter 3) was performed by conducting a correlation analysis between the QSBR model and the experimentally determined rates.

4.3 Experimental Procedures

4.3.1 Chemicals and Bacterial strains

Five test chemicals were selected from a previous chemical dataset (**Chapter 3**), considering two criteria: 1) selected chemicals cover the whole range of biodegradation rates and2) the availability of detailed aerobic biodegradation pathways for chemical.. Chemicals with known aerobic biodegradation pathways having the highest (phenol, toluene), lowest (2,4-dichlorophenol) and intermediate (4-chlorophenol, m-cresol) biodegradation rates were selected for this study (**Figure 4-1**).In general, the biodegradation of these chemicals can occur with one unique pathway (e.g., 2,4-dichlorophenol) and /or multiple pathways (e.g., 4-chlorophenol, phenol, toluene and m-cresol) depending upon the type of degrader and environmental conditions (i.e. aerobic and anaerobic) (Zylstra *et al.*, 1988; Gao *et al.*, 2010; Arora and Bae, 2014).

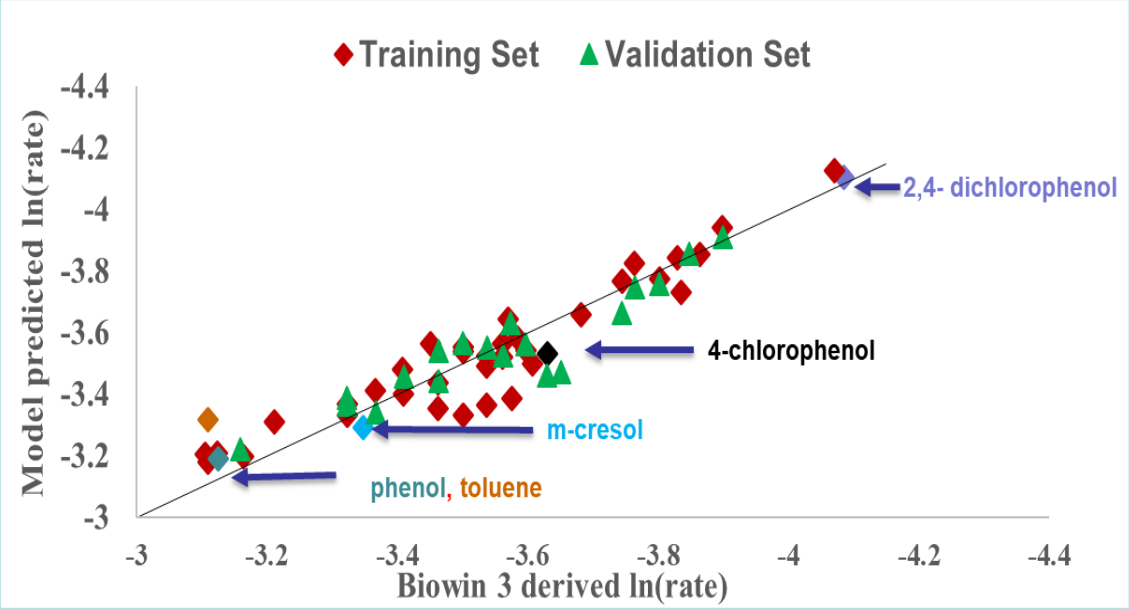


Figure 4-1 Plot of BIOWIN3 derived rates versus model calculated rates. Chemicals with known aerobic biodegradation pathways having the highest (phenol, toluene), lowest (2,4-dichlorophenol) and intermediate (4-chlorophenol, m-cresol) biodegradation rates were selected for biodegradation assays. The colour of data point matches the colour of the chemical names.

The details of probable aerobic biodegradation pathways of these chemicals are described elsewhere (Gao *et al.*, 2010). All chemicals were purchased from Sigma-Aldrich with > 99% purity (Dorset, United Kingdom). Pure cultures with metabolic capability to mineralize these chemicals were chosen as inoculum for the batch biodegradation experiments and were obtained from the Deutsche Sammlung von Mikroorganismen und Zellkulturen (DSMZ, Germany) **Table 4-1**. Upon arrival, all the bacterial strains were reactivated and stored in the prescribed media according to the supplier instructions. Growth studies of each bacterium were performed in nutrient broth medium (22 ^{o}C, 155 rpm).

Table 4-1 List of chemicals, their known degrader strains and their designation

Chemicals	Degrader	Culture collection
Phenol	*Pseudomonas putida* F1 (Reardon *et al.*, 2000)	DSMZ (6899)
Toluene	*Pseudomonas putida* F1 (Spain and Gibson, 1988; Zylstra *et al.*, 1988; Reardon *et al.*, 2000)	DSMZ (6899)
m-cresol	*Pseudomonas putida* F1 (Spain and Gibson, 1988)	DSMZ (6899)
4-chlorophenol	*Ralstonia sp*. strain RK1 (Steinle *et al.*, 1998)	DSMZ (11853)
2,4-dichlorophenol	Ralstonia sp. strain RK1 (Steinle *et al.*, 1998)	DSMZ (11853)

4.3.2 Experimental System

Experiments were carried out in reactors consisting of a 500 mL Duran Schott glass bottle, modified on the side and the top as shown in **Figure 4-2** (Final volume; 580 mL). The opening of the side tube was closed with a polytetrafluoroethylene **(**PTFE) coated rubber septum (13 mm butyl flange stopper; Capitol Scientific, Inc., Austin, USA), which was made tight using a screw cap. PTFE coating prevented the septum surface from adsorbing any chemicals. The top opening of the reactor was capped with Teflon Mininert caps (Sigma-Aldrich, United Kingdom). These type of caps have a push-button valve for sampling and offer leak-tight closures that minimize contact with the sample (Sigma-Aldrich, United Kingdom). Headspace gas was sampled using a gas tight 100 µL Hamilton syringe (Cole-Parmer, United Kingdom). Prior to running the biodegradation experiments, two control experiments were conducted in triplicate to confirm whether or not there was an apparent loss of CO_2 via diffusion through the gap between walls of the vials and the Mininert caps (method and results shown in **Appendix B, section 8**)

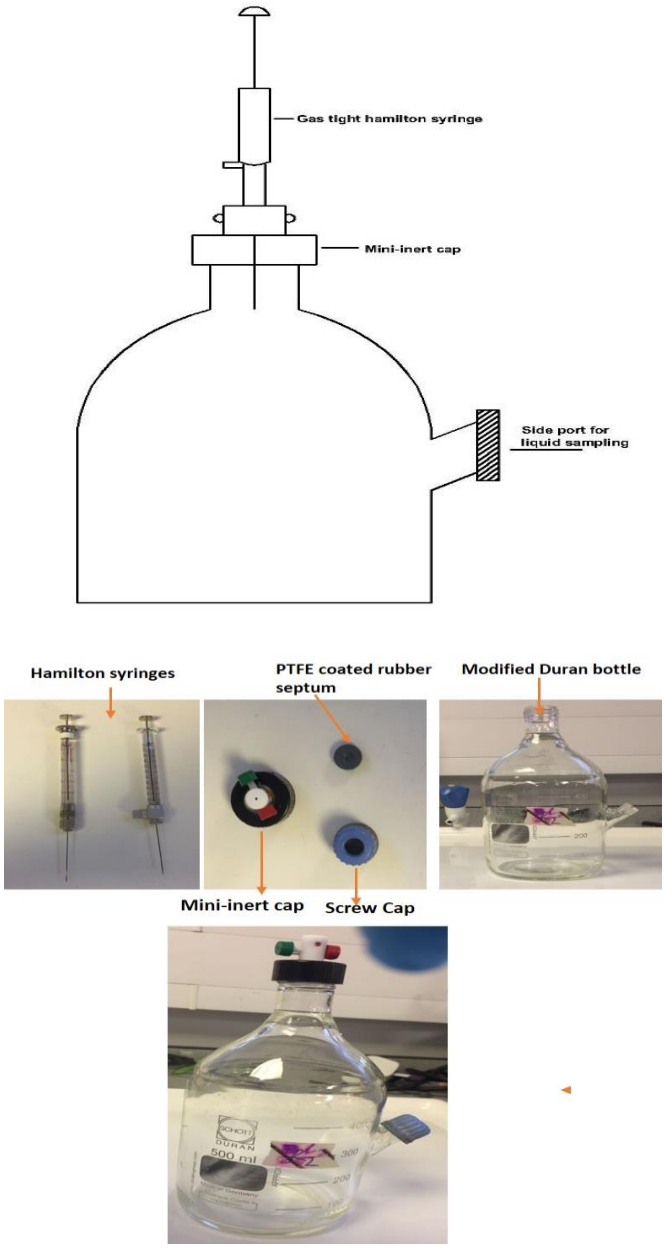


Figure 4-2 Illustration of the glass bottle reactor used to conduct biodegradation studies of test chemicals

4.3.3 Biodegradation assays

The biodegradation studies for each of the chemicals were carried out, one at a time, in the above batch reactors (**Figure 4-2**). The inoculum for each biodegradation experiment was taken during the exponential growth phase from a culture growing on nutrient broth. Approximately 10^7 cells/mL, quantified by flow cytometry (**Appendix B, section 4**), were inoculated aseptically into each sterile reactor bottle prefilled with 200 mL sterilized minimal microbial growth medium [6.78 g/L Na_2HPO_4, 3 g/L KH_2PO_4, 1 g/L NH_4Cl, 0.5 g/L NaCl] (Sigma – Aldrich, United Kingdom) containing 0.2 mL of trace element solution (**Appendix B, Section 1**). Prior to inoculation, bacteria were pelleted at 5000 x *g* for 10 minutes at room temperature and washed twice with phosphate buffered saline [PBS] (8 g/L NaCl, 0.2 g/L KCl, 1.15 g/L Na_2HPO_4, 0.2 g/L KH_2PO_4; pH 7.3). To maintain a constant pH in the reactor, 10

mM of sterile HEPES buffer (Sigma-Aldrich, United Kingdom) was added to the minimal media. 50 mg/L of test chemical was spiked into the reactor, which was then capped and wrapped in aluminium foil before being placed on a shaker table for two minutes to homogenize the reactor contents. All biodegradation assays were run in triplicate. An abiotic control was prepared in a similar way but without the inoculum. Reactors were incubated at 22 ^{o}C shaken at 155 rpm (Multitron Pro, INFORS HT, UK). 1.5 mL of liquid sample was collected from the side port using a sterile syringe (Syringe Discardit II 2 mL; VWR, United Kingdom) and needle (VWR, United Kingdom). 1mL of sample (except for the sample from the reactor with toluene) were clarified by centrifugation (8000 x *g*, 5 min) followed by filtration (0.2 μm PVDF syringe filter; VWR, United Kingdom). 0.5 mL of filtrate was mixed with an equivalent volume of methanol (Sigma-Aldrich, United Kingdom) and analysed to quantify the chemical removal by HPLC (**Table B 1, Appendix B**) (Shimadzu, UK). Toluene was extracted in hexane (Sigma-Aldrich, United Kingdom) using a liquid- liquid extraction method and the hexane extract was analysed by GC-FID (Agilent Technologies, Palo Alto, USA) according to the method described in **Appendix B (Section 3).** Further, headspace toluene in the reactor was also quantified over the duration of the experiment by GC-FID (**Appendix B, Section 3)**. The remaining cell solution was resuspended, 0.5 mL of which was analysed using a flow-cytometer (Becton Dickinson, California) to monitor the cell growth in the reactor (**Appendix B, Section 4**). The headspace CO_2 development during the experiment was quantified by GC-MS (**Appendix B, Section 5**).

4.3.4 Carbon mass balance

A total mass balance for carbon was carried out for each reactor; the initial and final total carbon content were determined. The total carbon in the reactor consisted of carbon present in the biomass, the headspace and the dissolved substrate, and headspace and dissolved CO_2 (including carbonate and bicarbonate). The carbon content of the biomass was estimated using the carbon content per bacterial cell (i.e. $9.4 * 10^{-14}$ g C/cell) mentioned elsewhere (Vrede *et al.*, 2002) according to **Equation 6.18** (**Appendix B Section 6**). The total CO_2 carbon content in the batch was estimated using **Equation 6.7 (Appendix B, section 6)** which is based on carbon dioxide and carbonic acid-base equilibria (Stumm and Morgan, 1970).
A theoretical carbon mass balance for each chemical mineralization was also performed. To achieve this, stoichiometric equations for microbial growth from chemical biodegradation were developed, with he assumption that NH_4^+ was the sole nitrogen source, using the method suggested by McCarty (2012). Briefly, bacterial growth involves two basic reactions: one for energy production and the other for cellular synthesis. The electron donor provides electrons

to the electron acceptor for energy production. A portion of its electrons (f_e) is initially transferred to the electron acceptor to provide energy for the conversion of another portion of electrons (f_s) into microbial cells. On a net yield basis, an assumption was made that 40% of the electron equivalent in electron donor substrate is used for synthesis ($f_s = 0.4$) (McCarty, 2012), while the remaining 60 % is used for energy ($f_e = 0.6$). Then, the overall energy and synthesis reactions were developed using the half reaction approach as described in **Appendix B, Section 9**.

4.3.5 Modelling biodegradation and estimation of kinetic parameters

As a minimum, a model of microbial processes must consider the mass balance of the active biomass and the primary substrate that limits the growth of the biomass. The relationship most frequently used to represent bacterial growth kinetics is the so-called Monod equation (Equation 1), which relates the specific growth rate of bacteria to the concentration of a rate limiting substrate as follows:

$$\mu = \mu_{max} \frac{S}{K_s + S} \quad [1]$$

Where μ = specific growth rate (s^{-1}), μ_{max} = maximum specific growth rate (s^{-1}), S = substrate concentration (moles C m^{-3}), K_s = substrate saturation constant (moles C m^{-3}) [i.e. substrate concentration at half μ_{max}]. The batch study (i.e. substrate removal, biomass growth and headspace CO_2 development with time) data was used to estimate the kinetic parameters (μ_{max} and K_s) for each chemical degraded. These parameters were estimated by fitting the experimental biodegradation results into a model simulating pollutant biodegradation according to Monod kinetics (Simkins and Alexander, 1984). The model evaluates the agreement between predictions and data for a range of μ_{max} and K_s parameter value combinations and identifies as the best-fit model parameters that give the minimum sum of squared residuals. The model was implemented in Matlab©, and the underpinning equations, and Matlab codes written by Dr. David Werner, and are provided in the **Appendix B (Section 6&7**).

The link between growth and substrate utilization has been described by Monod, who linearly related the yield coefficient (Y) to maximum specific growth rate of biomass (μ_{max}) and the maximum specific rate of substrate consumption or removal (q_{max}), as shown in Equation 2 (Pitter and Chudoba, 1990).

$$\mu_{max} = q_{max} \cdot Y \quad [2]$$

4.4 Results and Discussion

4.4.1 Batch biodegradation of chemicals

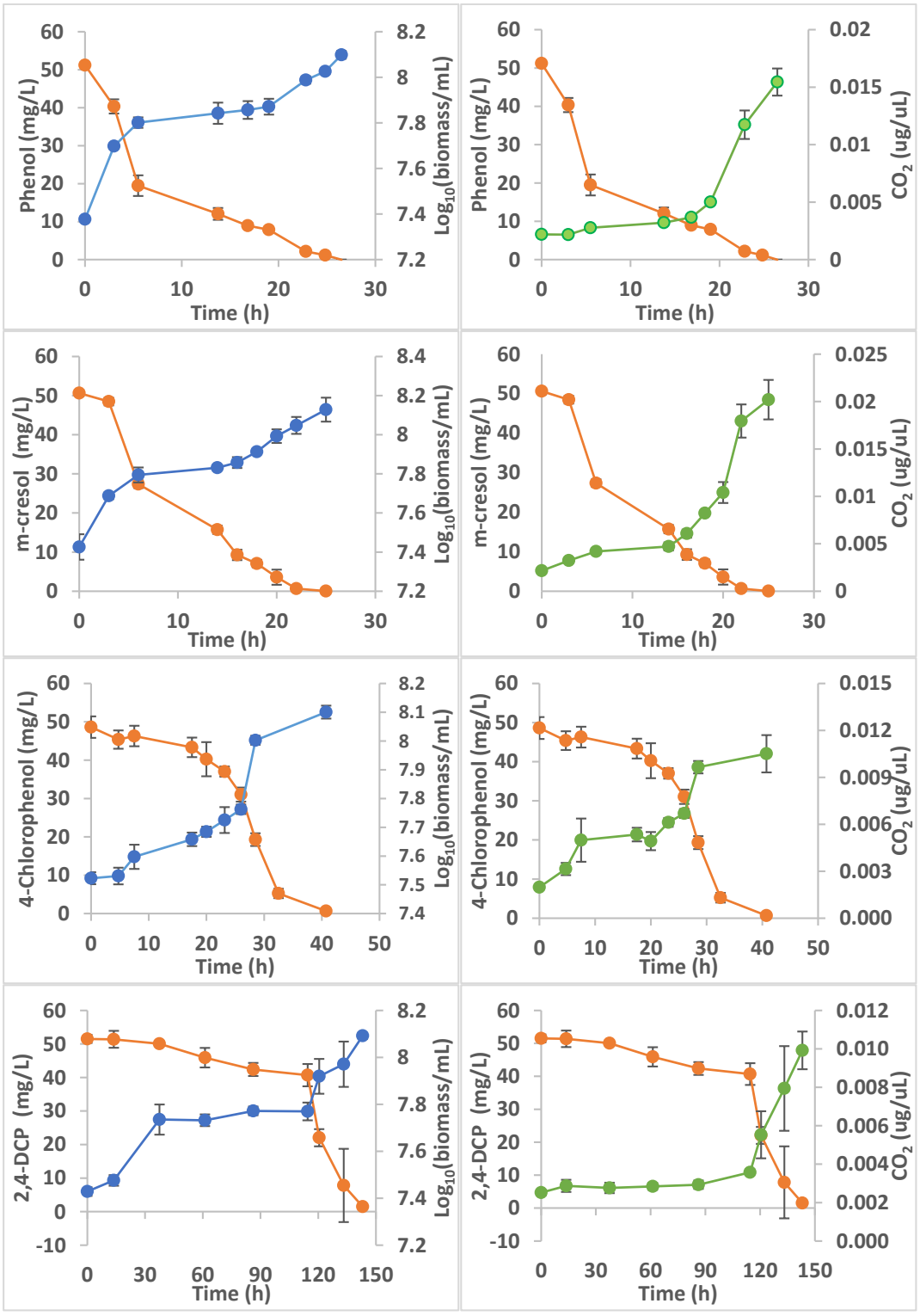
The concentration profiles of biomass, chemicals and CO_2 during aerobic biodegradation of five different chemicals at a fixed initial concentration (i.e. 50 mg/L) are shown in **Figure 4-3** and **Appendix B (section 13, Figure B 3)**. Biodegradation of phenol, toluene and m-cresol proceeded without any lag phase, while lag phases were observed for 4-chlorophenol and 2,4-dichlorophenol biodegradation (**Figure 4-2**). The biodegradation of test chemicals predominantly occurred during the exponential phase of microbial growth, as is commonly observed (Reardon *et al.*, 2000). This can be explained by the abundance of space, nutrients and desired conditions for growth (Pitter and Chudoba, 1990; Rolfe *et al.*, 2012) immediately following the lag phase, allowing microorganisms to grow at their maximum rate. All chemicals were biodegraded, as seen by a continual decrease in the chemical concentration and increase in the headspace CO_2 and biomass concentration. Furthermore, biomass growth in all the reactors was exactly mirrored by chemical, i.e. carbon substrate, consumption. The time required for complete biodegradation of each chemical differed **(Figure 4-3).** In all reactors, there was no significant difference between the initial and final pH values (two sample t-test, $p > 0.05$), since biological buffer was added to the growth medium. In the control experiments (results not shown) conducted in the absence of the inoculum, no change in the concentration of chemicals was observed, indicating abiotic processes were insignificant.

4.4.2 Biodegradation kinetics and parameters estimation

The different components associated with test chemical biodegradation in the batch reactors are presented in **Figure B 2 of Appendix B (section 10)** along with the Monod model prediction, which show how the best fit kinetic parameters for each chemical degradation were estimated. All the components are expressed in grams carbon per m^3. Examination of **Figure B 2 (Appendix B, section 10)** suggests that the Monod model provided a reasonable fit to the acquired experimental data, especially for substrate removal. However, the fit for CO_2 was not so good. This could be attributed to a CO_2 yield that varies as a function of time. The experimental data sometimes appears to show two phases of biodegradation, demonstrating more complexity in the biodegradation process than considered by Monod. This might be due the partial degradation of substrate initially to a metabolite or adsorption of substrate onto biomass, so there is a delay between the substrate degradation and the fully equivalent biomass and CO_2 production. On the other hand, the assumed partitioning

equilibrium between CO_2 in headspace and aqueous medium may not always be instantaneous as assumed by the model.

Good knowledge of initial values for Monod parameters is required, or convergence cannot be readily achieved (Zhang and Hughes, 2004). It should be noted that the range of fitting



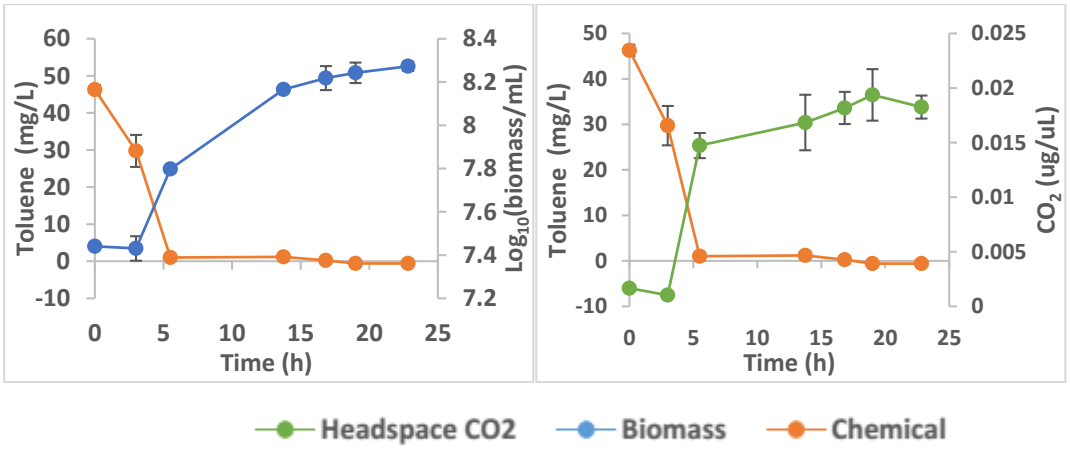


Figure 4-3 Chemical concentrations, biomass growth and headspace CO_2 concentration profiles obtained during biodegradation experiments with known degraders at room temperature for phenol, m-cresol, 4-chlorophenol and 2,4-dichlorophenol and toluene. Error bars represent the standard deviations for the triplicate experiments and might not be visible in some cases.

parameters (i.e. μ_{max} and K_s) that were used in the model was selected based on the initial fitting parameters calculated with experimental biodegradation results using a method mentioned elsewhere (iitd.vlab.co.in, 2012).

Table 4-2 presents the different observed kinetic and biological parameters for the biodegradation of each of the test chemicals in closed reactors. According to the Monod kinetic model (Equation 1, this chapter), the specific growth rate (μ) depends on the concentration of the limiting substrate. The test chemical concentration used in our study is comparatively higher than that routinely observed in the environment (Kowalczyk *et al.*, 2015b), and under the experimental conditions, the kinetics of degradation would be zero-order and therefore the substrate was removed with the maximum and highest rate (McCarty, 2012). The value of μ_{max} was slightly higher for *Pseudomonas putida* F1 when toluene (0.888 h^{-1}) was used as a substrate followed by m-cresol (0.755 h^{-1}) and phenol (0.666 h^{-1}). Whereas, the μ_{max} value for *Ralstonia* sp. strain RK1 was comparatively higher when 4-chlorophenol (0.266 h^{-1}) was used as substrate compared to 2,4-dichlorophenol (0.036 h^{-1}). The kinetics of enzymatic substrate uptake and microbial growth under limiting substrate conditions, the scenario commonly prevalent in the environment, is assessed by the ratio of μ_{max} and K_s (i.e. μ_{max}/K_s), also termed as specific affinity (Kovárová-Kovar and Egli, 1998). This ratio reflects the competitiveness of a microbial population to grow on a low concentration of limiting resource: the higher the value of μ_{max}/K_s, the better the microbial population grows on a limiting resource resulting in rapid substrate depletion (Healey, 1980a).

The yield coefficient for the biodegradation of each chemical was measured using the experimental results, which is the mass of biomass formed per unit mass of chemical

consumed (**Table 4-2)**. Yield coefficient data revealed that *Pseudomonas putida* F1 can produce comparitively higher amounts of biomass carbon per unit mass of toluene carbon than per unit masses of phenol and m-cresol carbon. This is clearly in agreement with the pattern observed in the case of aerobic biodegradation of toluene and phenol by *Pseudomonas putida* F1, where the higher yield coefficient was achieved for toluene at 30 ^{O}C (Reardon *et al.*, 2000). 4-chlorophenol and 2,4-DCP biodegradation data revealed that, for *Ralstonia baseliensis,* the efficiency was better for 2,4-dichlorophenol than for 4-chlorophenol. Regardless of substrate type, the yield varies according to growth rate, pH, temperature and the concentration of the excess substrate, and it can be used to predict the substrate concentration required to produce a certain biomass concentration (Hong, 1989; Pitter and Chudoba, 1990).

The test chemicals were employed as sole carbon sources in the reactor. μ_{max} and q_{max} for each chemical biodegradation are reported in **Table 4-2**. The pattern of maximum specific substrate removal rates for the studied chemicals was m-cresol > phenol > toluene> 4-chlorophenol > 2,4-dichlorophenol. However, q_{max} for toluene, m-cresol and phenol were relatively similar to each other and comparatively higher than for 4-chlorophenol and 2,4-dichlorophenol. The possible explanation for the observed differences in the substrate removal rate might be due to the difference in the type of substituent groups attached to the aromatic ring, as substituent groups are associated with electronic properties of chemicals and influence their biodegradability (Pitter and Chudoba, 1990). Additionally, the two organisms used in the biodegradation assays are different and may have different transport and metabolic efficiencies. However, studies have shown that, when glucose was used as a sole carbon source in mineral medium, *Pseudomonas putida* F1 and *Ralstonia eutropha* [95.7% homology with *Ralstonia* sp strain RK1 (Steinle *et al.*, 1998)] have maximum specific growth rate of 0.20 h^{-1} (Oliveira *et al.*, 2009) and 0.23 h^{-1} (Marangoni *et al.*, 2001), respectively, indicating the similar metabolic efficiency for glucose degradation. In aromatic chemicals, aerobic biotransformation generally starts with the addition of molecular oxygen to the aromatic ring (i.e. hydroxylation), by oxygenase or dioxygenase enzymes (Pitter and Chudoba, 1990; Peijnenburg, 1994), followed by aromatic ring cleavage. Hydroxylation of the aromatic ring with subsequent ring cleavage follows an electrophilic substitution mechanism and they are considered as rate determining steps (Pitter and Chudoba, 1990). The presence of substituent groups that increase the electron density of the aromatic ring accelerates the biodegradation process. Nitrogen and halogenic substituents strongly deplete the electron density of the aromatic ring, decreasing the rate of biodegradation as compared

to degradation of aromatic compounds with other substituents (e.g., OH, CHO, CH_3). The results in this study are in agreement with the results demonstrated by Alexander and Lustigman (1966), Martin et al (2017) and Pitter and Chudoba (1990). (Alexander and Lustigman, 1966) showed slower rates of biodegradation of mono- and di- substituted benzene by soil microorganisms with chloro-, nitro- and sulfonate- substituents, whereas an increased rate was recorded in the presence of hydroxyl and carboxyl groups. Similarly, (Martin *et al.*, 2017c) showed that, phenol substituent groups at position 3 (meta-position) have significant effects on the biodegradation potential of chemicals, where the sequential order of biodegradation by substituent groups was carboxylic acid (COOH) > hydroxyl (OH) > methyl (CH_3) > methoxy (CH_3O) > chloro (Cl) > nitro (NO_2) > bromo (Br) > fluoro (F). Likewise, a similar trend was observed in a study performed by (Pitter and Chudoba, 1990).

Table 4-2 Estimated Monod model parameters values for aerobic biodegradation of the test chemicals by known degraders

Chemicals	[a]Yield Coefficient [Y]	[b]K_s (mole C/m^3)	Maximum specific growth rate [c][μ_{max}]	μ_{max} /K_s	q_{max} [d]
Phenol	0.244	10	0.666	0.066	2.730
m-cresol	0.275	13.79	0.755	0.054	2.743
toluene	0.379	3.33	0.888	0.266	2.344
4-CP	0.352	7.5	0.266	0.035	0.758
2,4-DCP	0.397	2.14	0.036	0.016	0.090

The unit of Y, q_{max} and μ_{max} are (mole C biomass. mole C substrate^{-1}) and (mole C substrate. mole C biomass^{-1}.h^{-1}) and (mole C biomass. mole C substrate^{-1}.h^{-1}) respectively.

[a]Yield coefficient: mass of biomass C produced per unit mass of substrate C consumed during each chemical biodegradation

[b]K_s and [c]μ_{max} : best fit parameters obtained by fitting the biodegradation data with the Monod model

q_{max} [d]: Calculated with **Equation 2 (This chapter)**

4.4.3 Carbon mass balance and stoichiometry to assess uncertainty in the biodegradation experiments

Complete degradation of test chemicals was observed, although the time required for complete degradation differed for each test chemical. **Figure 4-4** presents the carbon mass balance of test chemical biodegradation in reactors. Mass balance showed an average recovery of 80.7 % of the initially applied carbon, ranging from 77.9 % to 81.5% recovery, depending on the test chemical. Variation between replicates for carbon mass balance recovery was typically low. The ability to account for 80 % of the introduced carbon in reactors provided an opportunity to determine the fate and partitioning of the test chemicals during the biodegradation process, either to inorganic carbon (CO_2) or its incorporation within the biomass. The production of metabolities is a significant phenomenon in the biodegradation of chemicals (Martin *et al.*, 2017b) and can potentially account for the

remaining fraction of missing carbon in the mass balance. Despite this, the phenomenon is probably insignificant in this study. Oxygen availability was not limiting so that complete mineralization of each chemical in the biodegradation tests was possible (**Table B 4, Section 12, Appendix B**). In addition, thermodynamic analysis suggested that the complete biodegradation of each chemical was feasible (**Appendix B, Section 9**). Furthermore, the absence of leaks from the reactor glass bottle **(Figure B 1, Section 8, Appendix B**), ensured that observations were not confounded by potential losses through volatility or leakage. However, uncertainty in the Henry constant, cell carbon content value used in the model, and biomass decay into dissolved organic carbon during the biodegradation reaction (the latter was not accounted for in the mass balance calculation), could have a significant effect on the final carbon content of the reactor. In this study, the cell carbon content value of $9.4 * 10^{-14}$ g C/cell (Vrede *et al.*, 2002) was used, which on average accounted for 43% of the carbon in the final amount of total carbon. In general, the elemental composition and cell volume of bacteria varies among species and is also influenced by environmental conditions (Trousselier *et al.*, 1997; Vrede *et al.*, 2002; Elazhari-Ali *et al.*, 2013). Previous studies have shown that the elemental content of bacterial cells varies across the different phases of growth: the highest carbon content was observed in exponentially growing cells, whereas the lowest was observed when the substrate starts to get limited (Vrede *et al.*, 2002). Carbon content per cell was therefore estimated in this study after complete degradation of the chemicals by assuming that all the substrate was converted into biomass and CO_2, that the phase partitioning and CO_2 speciation in the aqueous medium was accurately described, and that CO_2 was measured accurately (Refer **Appendix B, Section 14** for calculation). We observed different carbon contents per cell in *Pseudomonas putida* F1 for the degradation of different test chemicals: the highest values were observed during m-cresol ($1.84 * 10^{-13}$ g C/cell) degradation compared to phenol ($1.75 * 10^{-13}$ g C/cell) and toluene ($1.46 * 10^{-13}$ g C/cell) degradation. Similarly, the carbon content per cell for *Ralstonia* sp during 2,4-dichlorophenol and 4-chlorophenol degradation were $9.11 * 10^{-14}$ and $1.31 * 10^{-13}$ g C/cell, respectively. Therefore, mass balance data would potentially be more rigorous if carbon content of different strains can be determined experimentally for each chemical biodegradation. However, quantifying the cell carbon content of individual bactaria was outside the scope of this study. On average, 65 % of test chemical was mineralized to CO_2, ranging from 53.9% to 74.8%, depending on the test chemical (**Table 4-3**). Whereas, 35 % of test chemical carbon was incorporated into biomass (ranging 25.1% to 46%) (**Table 4-3**). Zhang and Hughes (2004) (Zhang and Hughes, 2004) studied the aerobic degradation of 2,4-

dinitrophenol (DNT) using DNT degrading mixed culture. These authors reported 25.3% of the substrate carbon was incorporated into bacterial cells, and the remaining 74.7% was evolved as CO_2 during complete biodegradation. In a similar study by Elazhari et al (2013), the author reported the average yield of 0.45 g of biomass carbon per gram of carbon (from biofuel) degraded (Elazhari-Ali *et al.*, 2013). These results suggest that large portion of carbon that are taken by cells are used for energy generation, rather than for incorporation into biomass. It should be noted that a major proportion of nutrients taken by bacteria are normally used for energy generation, simply to maintain cellular function (e.g., maintain membrane potential, maintain an osmotic pressure across bacterial membrane, synthesizing building blocks, take up or excrete some compounds against the concentation gradients) (Hu, 2017).

In this study, a stoichiometric equation was developed for the biodegradation of each test chemical and the theoretical yield (biomass and CO_2) was compared with experimentally observed results (**Appendix B, section 9**). **Table 4-3** shows the theoretically estimated and experimentally observed CO_2 and biomass amount for each test chemical. The average theoretical amount of biomass formation and CO_2 production were 0.46 mole of biomass C/mole of substrate carbon and 0.54 mole of CO_2 C/mole of substrate carbon respectively, ranging from 0.4 to 0.514 mole of biomass C/mole of substrate carbon and 0.486 to 0.6 mole of CO_2 C/mole of substrate carbon, depending on the test chemicals. The average experimental amount for biomass and CO_2 formation were 0.3015 mole biomass C/ mole of substrate carbon (range: 0.218 – 0.411 mole of biomass C/mole of substrate carbon) and 0.557 mole of CO_2 C/ mole of substrate carbon (0.481 – 0.649 mole of CO_2 C/mole of substrate carbon) respectively, depending on the test chemical. Thus, the experimentally observed biomass formation was typically lower than the theoretically estimated biomass formation except for 2,4-DCP biodegradation, although the range of values measured encompasses the range of theoretical predictions. In has to be noted that, the theoretically estimated yield (i.e. true yield) is based on the material balance equation between cells, substrate and the products and does not take into account biomass loss due to cell decay or endogenous respiration, and also the energy consumed by cells for the maintenance, and is therefore always higher than net yield (i.e. experimentally observed yield). An inability to estimate the rate of biomass decay or self respiration during the mineralization process is a limitation of the experimental system approach used in this study.

Table 4-3 Comparison of theoretically estimated biomass and CO_2 amount for each test chemical biodegradation with the experimentally observed amounts. A theoretical carbon mass balance for each chemical mineralization was also performed with stoichiometric equations for microbial growth using the method suggested by McCarty (McCarty, 2012) [Appendix B, Section 9]

	Theoretical Yield		Experimental Yield	
	[a]CO_2	[b]Biomass	[a]CO_2	[b]Biomass
Phenol	0.533	0.467	0.6 ± 0.055	0.251 ± 0.012
m-Cresol	0.514	0.486	0.649 ± 0.075	0.218 ± 0.013
Toluene	0.486	0.514	0.539 ± 0.037	0.326 ± 0.019
4-chlorophenol	0.567	0.434	0.519 ± 0.039	0.302 ± 0.021
2,4-dichlorophenol	0.600	0.400	0.481 ± 0.043	0.411 ± 0.017

[a] moles of CO_2 carbon produced per mole of substrate carbon
[b] moles of biomass carbon produced per mole of substrate carbon

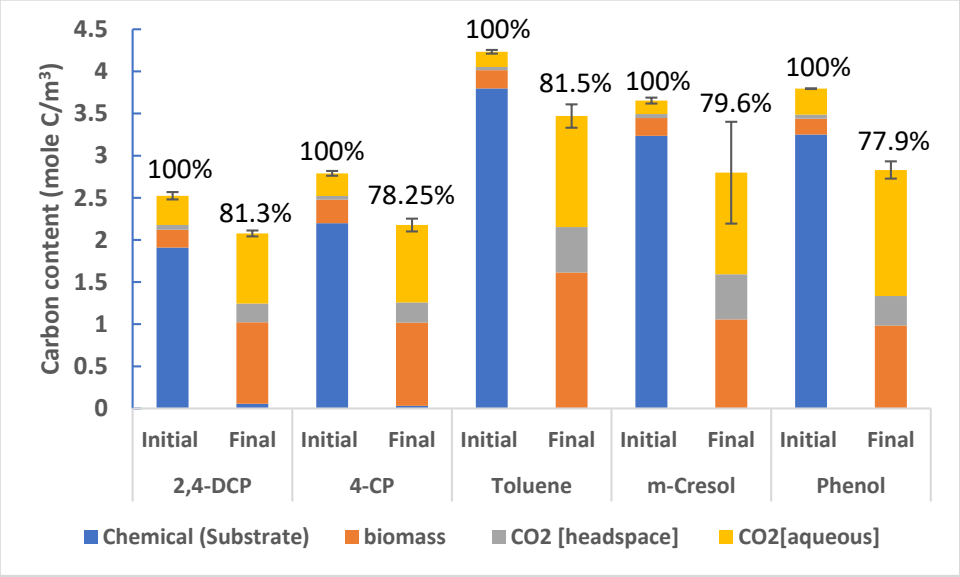


Figure 4-4 Clustered stacked column chart summarizing the mass balance result for each test chemical biodegradation in reactors. Error bars represent the standard deviations for the total carbon content triplicate experiments.

Table 4-4 Summary table for correlation (Pearson correlation) values. The reported values are correlations between the observed biodegradation rates of test chemicals and the individual molecular descriptors from QSBR model developed previously.

	[a]Mor08u	[b]nArX	[c]tot.energy	[d]E_{homo}	[e]E_{lomo}	[f]Elec. energy	[g]MW	[h]vdw V	[i]MR	[j]Log P	[k]σ	[l]$_vX^1$	[m]$_vX^2$
R^2	-0.272	0.920	0.589	0.775	0.835	0.618	0.806	0.546	0.540	0.416	0.689	0.623	0.613
Std. Error	1.385	0.348	0.788	0.582	0.500	0.759	0.541	0.828	0.833	0.939	0.685	0.755	0.764
p-value	0.730	0.006	0.081	0.031	0.019	0.072	0.025	0.095	0.097	0.145	0.052	0.070	0.073

[a]: un-weighted descriptor with scattering parameter (s) = 7 $Å^{-1}$ (3D Molecular Representations of Structure based on Electron diffraction); [b]: Number of halogen on aromatic rings (Functional group counts); [c]: Total energy (ToE) of molecule; [d]: Energy of highest occupied molecular orbital; [e]: Energy of lowest occupied molecular orbital; [f]: Electrical energy of molecule; [g]: Molecular weight; [h]: van der Waal Volume of molecule; [i]: Molar Refractivity; [j]: n-octanol and water partition coefficient; [k]: Substituent (hammett) constant; [l]: 1st order valence connectivity index; [m]: 2nd order valence connectivity index

4.4.4 Verification and calibration of a QSBR Model

The main objectives of this study was to verify a previously developed QSBR model (Chapter 3) for mono-aromatic chemicals with experimental biodegradation rates of selected test chemicals. In the environment, the rate of chemical biodegradation is uncertain and difficult to reproduce as it is influenced by spatial and temporal variability of combination of different factors (i.e. abundance and activity of degrading microorganism, environmental conditions, structure and concentration of chemicals) (Pavan and Worth, 2008). The biodegradation rates used for the derivation of the QSBR model are the rates of test chemical degradation in the aquatic environment (i.e. using environmental inocula). In contrast, the different parameters influencing biodegradation can be more controlled in laboratory biodegradation tests. Particularly, biodegradation data obtained from simplified biodegradation system (i.e. batch biodegradation assays, where the chemical act as the only source of carbon and energy) are easily interpretable and comparitavely reproducible as compared to biodegradation data obtained from mixed culture assays, and might be suitable towards calibrating theoretical QSBR models. However, it can't be ignored that several factors can effect degradation of chemicals in pure culture studies, which could ultimately have influence on the estimated bidegradation rate and thus on calibration of QSBR models. For example, culturing degraders in the nutrient broth media prior inoculating them into the mineral medium with chemical might contribute in the possible loss of biodegradaing capacity of degrader, and therofore, lag phase can be observed during the biodegradation assays. In addition, enzymes needed for catalyzing degradation reactions are to be induced.

Direct comparison of the biodegradation rates used in the previous QSBR study and experimentally determined rates from this study for selected test chemicals is a misleading metric for verification of the previous QSBR model, because the previous rates were derived from the BIOWIN3 model (pass/fail model for biodegradibility), and transformed into rates that do not reflect those measured experimentally in this study. For example, the BIOWIN3 derived rates are based on environmental, mixed communities, not pure cultures.

One way in which the principles of the previously derived QSBR model could be verified and calibrated is through correlation analysis between the molecular descriptors from the previous QSBR model **Table 4-4** and the experimentally determined biodegradation rates in this study. In addition, the comparison of patterns of biodegradation rates of test chemicals in the previous QSBR model study and this experimental study could further validate the principle and suitability of the previous model.

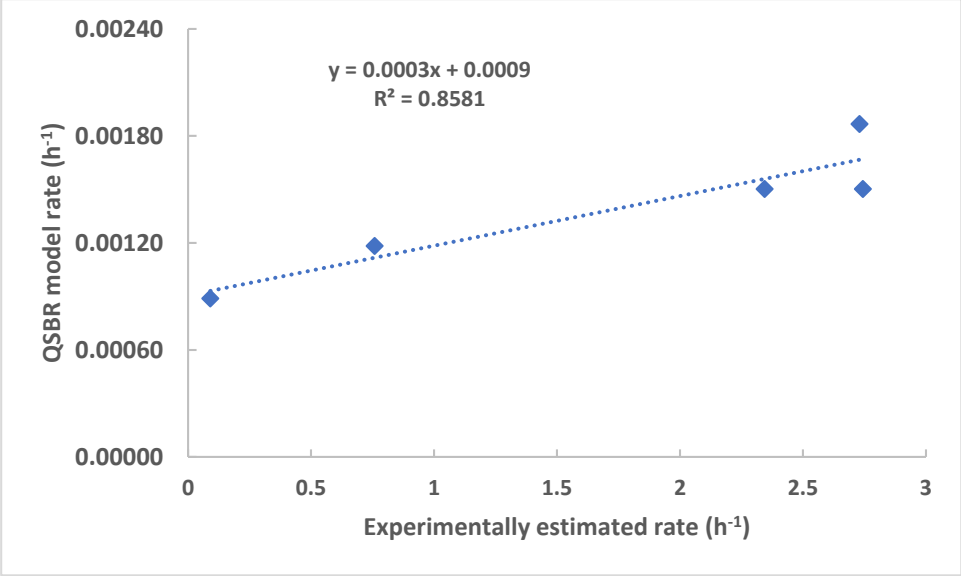


Figure 4-5 Relationship between the experimentally determined biodegradation rates and the biodegradation rates used in previously developed QSBR models for five test chemicals. The QSBR model rates are approximately 1000 folds lower than experimentally estimated rates. The is mostly attributed to differences in inocula types and concentration, initial chemical concentration, environmental conditions (i.e. temperature, medium constituents, pH, dissolved oxygen concentration) used for conducting biodegradation assays.

A strong and significant correlation ($p < 0.05$, $r^2 = 0.85$), along with similar patterns of biodegradation rates, were observed between the experimentally determined rates and those rates used in the previously developed QSBR model for five of the test chemicals **(Figure 4-5**). To test the statistical association between experimentally determined biodegradation rates and each of the molecular descriptors, univariate regression analysis was performed (**Table 4-4 and Table B 3 [Appendix B, Section 11] for details**). In short, reported R^2 values reveal a significant correlation between q_{max} and the majority of the descriptors listed for monoaromatic chemicals included in the previous study. Because of the small number of samples, it was not feasible to perform multivariate regression using combinations of two or three descriptors.

It is interesting to note that molecular descriptors given in **Table 4-4 r**epresent a variety of physicochemical, structural and quantum mechanical properties of chemicals that have previously shown an association with biodegradation rates (Okey and Stensel, 1993; Okey and Stensel, 1996; Yang *et al.*, 2006b). The observed correlation between biodegradation rates and these descriptors suggest that different factors play an important role in the biodegradation of mono-aromatic chemicals that may offer some explanation for the observed differences. More specifically, enzyme binding, chemical transformation, kinetics and thermodynamic factors are pivotal during the biodegradation of chemicals and the importance of these factors is discussed elsewhere (Parsons and Govers, 1990b; Wammer and Peters, 2005).

A lack of reliable and robust experimental data have fundamentally undermined the use of QSBR models in regulatory capacity (Martin *et al.*, 2017c). The biodegradation test method developed in this study enabled the generation of reliable biodegradation rates that were used to verify and validate the principles behind a previously developed *in silico* QSBR. Overall, these results provide evidence that the principles used in developing the previous QSBR model were robust and appropriate in predicting the relative biodegradibility of chemicals within the same applicability domain.

4.5 Conclusions

A lack of reliable and robust experimental data has fundamentally undermined the use of QSBR models in a regulatory setting. The biodegradation test method developed in this study enables the generation of reliable biodegradation rates that can then be used to validate and calibrate the principles behind *in silico* QSBRs. In addition, a rigorous carbon mass balance model and stoichiometric equations for chemical biodegradation can enhance the quality assurance of measured biodegradation rates and address the uncertainty associated with the experimental determination of biodegradation rates.

The previous QSBR model was quantitatively validated using the biodegradation rates derived from pure culture experiments that demonstrated that the same rank prioritisation of biodegradation rates existed even if the absolute rates were different. The next step would then be to test the biodegradation of the same set of chemicals with a mixed microbial community, for example activated sludge. I believe that a traditional biodegradation test for a few representative chemicals among the set of chemicals used in QSBR model development can be performed with inocula simulating the environmental compartment of interest, and the estimated experimental biodegradation rate can be ultimately used to verify and calibrate such QSBR models. The simple but rigorous experimental and theoretical approach used in this study could form the basis for defining rate boundaries for calibration against the QSBR data.

Chapter 5

Development and application of molecular tools to study the biodegradation of mono-aromatic chemicals in activated sludge

Chapter 5. Development and application of molecular tools to study the biodegradation of mono-aromatic chemicals in activated sludge

5.1 Abstract

This work focuses on accessing the biodegradation potential of natural mixed communities for phenol and 2,4- dichlorophenol (2,4-DCP) by identifying and quantifying putative chemical degraders in microbial communities, with the use of key catabolic genes involved in the chemical degradation pathway. Degenerate primer-sets were designed or adopted from the literature to target and quantify these genes and their transcripts using quantitative real-time polymerase chain reaction (qPCR) and Reverse Transcription-qPCR – RT-qPCR. This approach, combined with traditional laboratory biodegradation studies, enabled the assessment of the biodegradative potential of environmental inocula and characterize shifts in functional genes and their expression during the biodegradation of phenol and 2,4-dichlorophenol (2,4-DCP). The results indicated that the test chemicals are eliminated biologically in activated sludge as a result of both the enrichment of potential degraders and expression of relevant genes in potential degraders present in indigenous population. Although phylogenetically diverse groups of bacteria are present in an environment, the presence or absence of bacteria possessing the catabolic ability to degrade specific chemicals determines the degradation potential of an environment for given non-readily biodegradable chemicals. Application of DNA and RNA based molecular techniques in biodegradation studies will improve our understanding of the composition and phylogeny of putative chemical degraders of mixed microbial community in the environment.

5.2 Introduction

The presence and activity of potential degraders in microbial communities is one of the major factors that influence the degradation of aromatic chemicals in the environment (Futamata *et al.*, 2001; Pavan and Worth, 2008; Tuan *et al.*, 2011; Kowalczyk *et al.*, 2015b; Peng *et al.*, 2015). Yet its measurement receives little practical attention in the investigation of specific chemical degradation by mixed microbial communities in the context of regulatory biodegradation tests (Kowalczyk *et al.*, 2015b). The activity of degraders in the environment is determined by the species present in the microbial communities, their relative population densities, the induction and expression of key catabolic genes, and their capability to grow on a chemical once exposed to it (Pavan and Worth, 2008). Therefore, application of molecular tools to identify and quantify potential degraders and their activities can be a useful metric to estimate the biodegradation potential of environmental inocula for specific aromatic

chemicals. In addition, it helps in understanding the link between active microbial communities and the processes they mediate.

Recent developments in molecular based methods have led to rapid and accurate strategies for monitoring, discovering and identifying active degraders and their catabolic genes involved in the degradation of aromatic chemicals (Widada *et al.*, 2002; Cébron *et al.*, 2008; Meynet *et al.*, 2015). Various studies have been conducted to quantify the bacteria involved in aromatic chemical degradation through quantitative PCR, using degenerate PCR primers targeting the functional aromatic chemical degrading genes (Sei *et al.*, 1999; Mesarch *et al.*, 2000; Baldwin *et al.*, 2003; Cébron *et al.*, 2008; Lillis *et al.*, 2010; Kowalczyk *et al.*, 2015a). More recently, the application of RNA-based techniques in environmental biodegradation studies have also been advocated (Widada *et al.*, 2002; Devers *et al.*, 2004; Helbling *et al.*, 2012) as they are able to distinguish the metabolically active portion of the microbial population in a system.

In the environment, there is often a broad range of potential degraders from phylogenetically diverse groups, which are capable of catabolizing aromatic chemicals via either metabolism or co-metabolism, under both aerobic and anaerobic conditions (Zylstra *et al.*, 1988; Bamforth and Singleton, 2005; Gao *et al.*, 2010; Arora and Bae, 2014; Meynet *et al.*, 2015). The aerobic metabolism of aromatic chemicals commonly initiates with the incorporation of molecular oxygen into the aromatic ring by monooxygenase or dioxygenase enzymes, forming dihydrodiol (catechol moieties) with subsequent aromatic ring cleavage (Pitter and Chudoba, 1990; Cerniglia, 1993; Reardon *et al.*, 2000; Iwai *et al.*, 2011). These reaction steps are often considered the rate-limiting step in biodegradation pathways, with genes involved in these reactions considered as key catabolic genes (Widada *et al.*, 2002; Cébron *et al.*, 2008; Lillis *et al.*, 2010). Primer-sets targeting key catabolic genes, designed either with amino acid or nucleic acid sequences, can be used as biological markers to assess the biodegradation potential of an environmental system for a given chemical (Mesarch *et al.*, 2000; Hall, 2011). Phylogenies based on primary nucleic acid sequence can be used as an alternative approach to using amino acid phylogenies for designing primer-sets which targets genes either for diversity or abundance studies (Meynet *et al.*, 2015). This approach reduces the degeneracies created from the codon redundancy when amino acid sequences are translated to nucleotide sequences.

The main aims of this study were;

(1) To design degenerate primer-sets able to detect and quantify each of the identified catabolic genes involved in key steps of mono-aromatic chemical biodegradation in environmental systems

(2) To optimize the primer-sets to investigate chemical degradation by mixed microbial communities (i.e. activated sludge) for mono-aromatic chemicals having different biodegradabilities.

Phenol and 2,4-dichlorophenol (2,4-DCP) were selected as model chemicals, as they are classified as readily biodegradable and not readily biodegradable chemicals (EPA, 2011) respectively, and were also included in our previous chemical dataset for QSBR model development (Chapter 3). Molecular microbial assays (qPCR and RT-qPCR) were developed and optimized for each primer-set to assess and quantify the activity of potential putative chemical degraders during biodegradation assays of test chemicals, namely phenol and 2,4-DCP in activated sludge.

5.3 Material and Methods

5.3.1 Selection of catabolic genes and reference microorganisms

Genes synthesising key catabolic enzymes for the degradation of phenol and 2,4 DCP were identified using the Pathway Prediction System database (Gao *et al.*, 2010) and are shown in **Table 5-1.** The reference microorganisms possessing the catabolic genes of interest were selected from the literature **(Table 5-1**) (Powlowski and Shingler, 1994; Leveau *et al.*, 1999).

Table 5-1 Key catabolic genes involved in phenol and 2,4-DCP aerobic biodegradation pathways and bacterial referene strains used in this study

	Compounds	Catabolic Enzymes	Reference microorganism
1	Phenol	Phenol hydroxylase (PH, [*DmpN*]) Catechol-2,3-diooxygenase (C23D [*DmpB*])	*Pseudomonas putida* CF600 *(Powlowski and Shingler, 1994)*
2	2,4-DCP	2,4-DCP hydroxylase (*tfdB*) 3,5-dichlorocatechol-1,2-dioxygenase (*tfdC*)	*Ralstonia eutropha* JMP134 *(Leveau et al., 1999)*

5.3.2 Phylogenetic Analysis

Baldwin et al (Baldwin *et al.*, 2003) and Futamata *et al* (Futamata *et al.*, 2001) previously designed the degenerate primer pairs for the phenol hydroxylase gene, however the sequence library was created without eliminating duplicate sequences, which was avoided in this study. Duplicate sequences do not provide any additional information to the tree, while increase the computational time and clutter the appearance of phylogenetic tree. The nucleic acid sequences of the selected genes (i.e. those for PH [*dmpN*], C23D [*dmpB*] and *tfdB*) from reference microorganisms (**Table 5-1**) were used as a reference sequences for the construction

of a sequence library. The GeneBank database, using the BLAST (Basic Local Alignment Search Tool) algorithm, was used to search for homologous sequences showing high identity (> 85%), query coverage ≥ 60% and E (expectation) value < 10^{-3}. Duplicate (identical) sequences were eliminated by computing the pairwise distance in MEGA 7 (Tamura *et al.*, 2011), and final library with relevant sequences was created (**Table C 1, Appendix C for details**). Clustal W (Larkin *et al.*, 2007) was used to perform pairwise and multiple alignment for relevant sequences, followed by construction of a phylogenetic tree in MEGA7 (Tamura *et al.*, 2011) by using the neighbour-joining method (**Figure 5-1, A B and C**).

5.3.3 Primer design

Primer-sets for chlorocatechol-1,2-dioxygenase (*tfdC*) gene were adapted from elsewhere (Lillis *et al.*, 2010), whereas other primers were manually designed here (**Table 5-2**), using the most conserved regions identified by the alignment of the corresponding nucleotide sequences. The designed primers were limited to length of 21 -26 oligonucleotides, annealing temperature between 50 and 61 ^{o}C, GC content ≤ 60 %, avoiding regions of self-complementarity, and preventing, where possible excessive degeneracies. The specificity of the chosen primer-sets (for both reference DNA and total DNA from activated sludge inocula) was checked and confirmed theoretically against the GeneBank database and the target microorganisms (i.e. genes) in the environment amplified by the primer sets are shown by using the phylogenetic tree in **Appendix C (Figure C 5, C 6, C 7 and C 8).** While, sequencing the PCR products (with both reference DNA and total DNA from activated sludge inocula) can further confirm the primer specificity, however was not considered in this study.

5.3.4 Bacterial strain and culture conditions

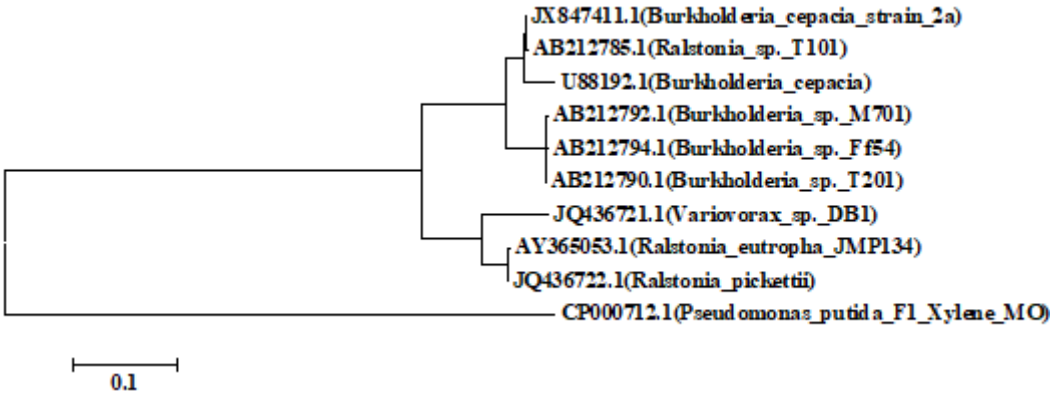
In order to test the specificity of the designed and literature-selected primer-sets, cultured reference bacterial strains were selected and obtained from culture collection (*Ralstonia eutropha* JMP134 from DSMZ*)* or laboratory (*Pseudomonas putida* CF600 from Prof. Dr. Vicky Shingler, Department of Molecular Biology, Umeå University). The organisms were grown either in a minimal medium or broth medium at a suitable temperature in a shaking incubator at 155 rpm. Respective solid media were prepared by the addition of 15 g/L agar. Details on bacterial strains and growth conditions are reported in **Table C 2 (Appendix C)**.

5.3.5 DNA extraction

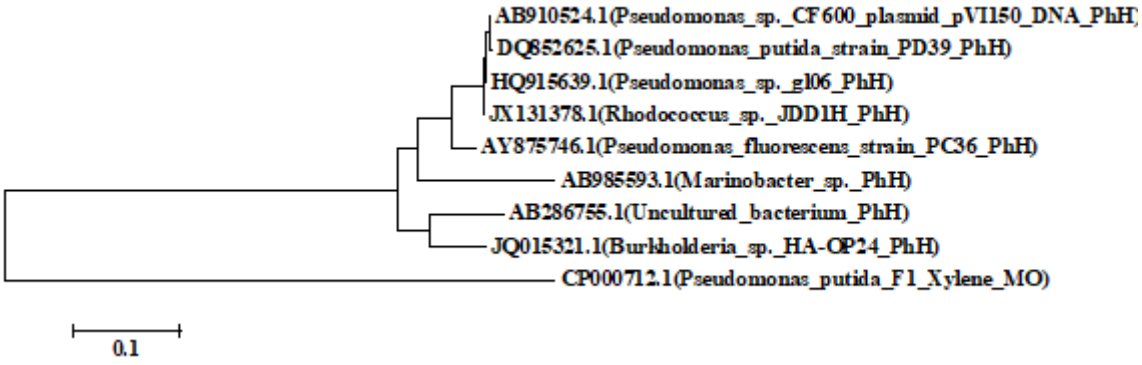
The identified catabolic genes involved in the degradation pathways of phenol and 2,4-DCP are encoded in the plasmid of the reference microorganisms (Powlowski and Shingler, 1994; Leveau *et al.*, 1999). The extraction of plasmids from reference organism was performed using the PureYield™ Plasmid Miniprep System as per the manufacturer's instructions

(Promega, United Kingdom). The concentration of the plasmid extracts was determined using a Qubit® dsDNA HS Assay Kit (Invitrogen, Life Technologies, Paisley, UK).
The extraction of total DNA from activated sludge was conducted using the FastDNA Spin Kit for Soil according to the manufacturer's instructions (MPBiomedicals, Santa Ana, CA, USA). Whereas, the DNA from non-target organisms (see below Section 5.3.7.2) was conducted using the ChargeSwitch™ gDNA Mini Bacteria Kit (Invitrogen, Life Technologies, Paisley, UK). The concentration of the DNA extracts was determined using a Qubit® dsDNA HS Assay Kit (Invitrogen, Life Technologies, Paisley, UK).
The principle of each kit for DNA and RNA (see below) extraction are briefly summarized in **Appendix C.10**.

[A]



[B]



[C]

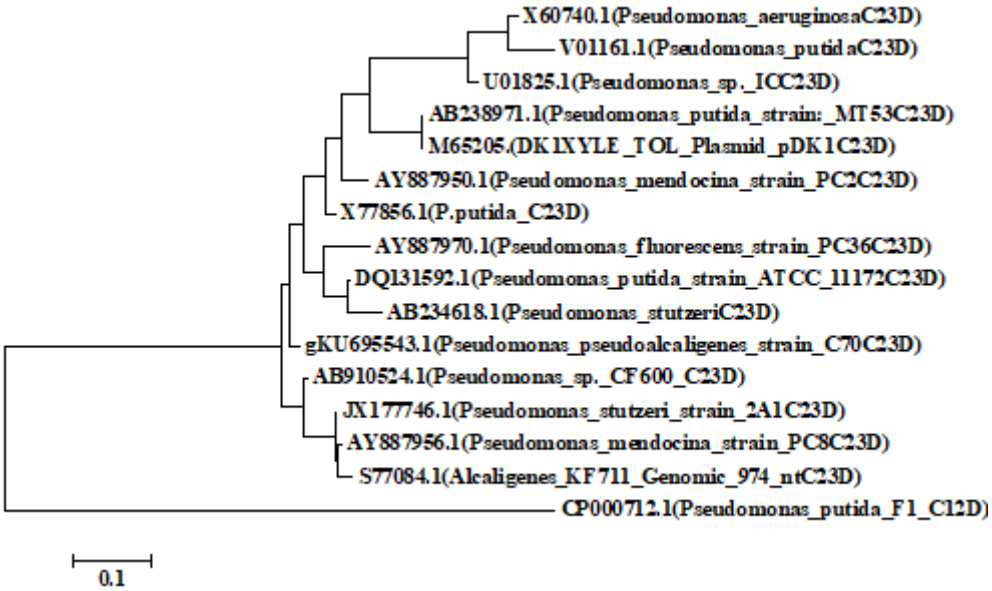


Figure 5-1 Unrooted phylogenetic neighbour-joining tree for; [A] 2,4-dichlorophenol hydroxylase, [B] Phenol hydroxylase, and [C] catechol-2,3-dioxygenase genes based on nucleotide sequences taken from the GeneBank. The bacterial sequences (except final one from each group) that are grouped (A, B and C) have been used for the final nucleic acid sequence alignment for tfdB, PH and C23D gene primer design, respectively. The final microorganism in each group was used as a non-target organism. Nucleic acid sequences were aligned using Clustal W (Larkin *et al.*, 2007) incorporated in MEGA7(Tamura *et al.*, 2011). The phylogenetic tree is constructed by using neighbour joining method with MEGA7 software (Tamura *et al.*, 2011). The bar represents 10% sequence divergence.

Table 5-2 List of primers used for the amplification of gene fragments coding for PH, C23D, *tfdC* and *tfdB* using PCR and real-time PCR

Gene	Primer	Position in Model Organisms	Sequence* 5' – 3'	Expected amplicon size (bp)	GC%	Main target genes	References
multicomponent Phenol hydroxylase (PH)	F	4303-4323	CCATGAGCCA**Y**TACAACAAGC	174	47.6	*dmpN*, *LmPH*	This study
	R	4458-4478	CA**R**CAGGTTGGTCAGCACGTA		52.4		
Catechol-2,3-dioxygenase (C23D)	F	7821-7841	CAGTTCCTCAG**Y**CTGTCGACC	205	60	*dmpB*, *pheB*, *PHB*, *phlH*, *nahH*, *bphE*, *xylE*	This study
	R	8004-8025	CGAAGAAGTAGATGGTCTTGCC		50		
2,4-DCP hydroxylase (*tfdB*)	F	27447 - 27468	G**Y**TCCTCGT**Y**GTAGGT**S**GC**K**AGCA	199	60.4	*tfdB*	This study
	R	27621 - 27646	GTGGCGTACATGTC**R**TTCAC**S**GTCCA		55.8		
chlorocatechol-1,2-dioxygenase (*tfdC*)	F	30524 - 30543	CCGCC**Y**TCGAAGTAGTA**Y**TGGGT	274	66.7	*tfdC*, *tcbC*, *clcA*	Lillis et al. (2010) (Lillis *et al.*, 2010)
	R	30782 - 30798	GTGTGGCA**Y**TCGACGCCGGA**Y**		57.1		

*Bases in bold correspond to degenerate bases in IUPAC codes (Y = T, C; R = G, A; K = G, T; S = G, C)

5.3.6 RNA Extraction

The total RNA was extracted from the actively (exponentially) growing culture in M9 Minimal Medium (Sigma-Aldrich, UK) with the relevant selected chemical acting as a sole carbon source (Phenol, 50 mg/L and 2,4-DCP, 5 mg/L), using the ISOLATE II RNA Mini Kit (Bioline, United Kingdom), following manufacturer's instructions.

Additional cleaning steps during RNA extraction were required for activated sludge (previously stored at -80 ^{o}C) due to the presence of a complex matrix, therefore RNA was isolated using the RNeasy PowerSoil Total RNA Kit according to the manufacturer's instructions (Qiagen, Crawley,UK).

DNA was removed from the extracts by treating RNA with DNase I according to the manufacturer's instruction (Invitrogen, Life Technologies, Paisley, UK) and by further clean-up using RNeasy MinElute Cleanup Kit (Qiagen, Crawley,UK). The concentration and quality of the RNA extracts was determined using Nanodrop 1000 spectrophotometer (Thermo Scientific) and reported in **Table C 3 (Appendix C)**.

5.3.7 PCR Analysis

5.3.7.1 Annealing Temperature optimization

The annealing temperature of each primer-set was optimized using plasmid DNA isolated from the respective reference microorganism. A gradient of annealing temperatures around the theoretically calculated annealing temperature, $T_a \pm 4$ ^{o}C, was used for each primer-set. The PCR reactions were performed in a total volume of 25 µL, containing 0.5 µM of each primer, 1 µl DNA template (or sterile molecular grade water, for the negative control and 1 µl DNA from non-target organism to check non-specific amplification [details provided in section **5.3.7.2**]), 0.2 mM of dNTPs, 1 U Taq DNA polymerase, (Invitrogen, Life Technologies, Paisley, UK), 1.5 mM $MgCl_2$ and 2.5 ul of 10 x PCR buffer (minus Mg). The following thermocycler program was used for the amplifications: a denaturation cycle of 94^{o}C for 3 minutes followed by 30 cycles of denaturation at 94^{o}C for 45 seconds, 30 seconds at annealing temperature, $T_a \pm 5$ ^{o}C and elongation at 72 ^{o}C for 1 minute, with a final elongation step at 72 ^{o}C for 10 minutes. Until further use, the PCR products were stored at -20^{o}C. The amplification of the correct size of gene fragments and the purity of PCR products were confirmed by routinely performing electrophoresis, run at 100 V for approximately 1 hour, using 7 µL of product on 1% agarose gels containing 0.1 µL/mL concentration of Nancy 520 DNA stain (Sigma-Aldrich, UK), and in 1 x Tris-acetate-EDTA buffer. Agarose gels were viewed in a UV-P gel documentation system (Genetic Research Instrumentation,

Dunmow, UK). The temperature at which a strong band was observed was considered as the optimum annealing temperature (**Figure C 2, Section1, Appendix C.6**).

5.3.7.2 Specificity

The specificity of each primer-set was checked by PCR against the reference microorganism and the non-target microorganism (to check for non-specific amplification) at the determined optimum annealing temperature. Furthermore, non-specific amplification in the total environmental DNA during PCR and qPCR can be furthered confirmed, respectively by sequencing the PCR amplified product and analysing the melt curve peak. However, former was not considered in this study. The selected non-target organism can be an organism that has similar genes (able to degrader specific test chemicals), but may mismatch with the nucleotide sequence of selected genes from reference organisms or can be similar genes, but different function. *Pseudomonas putida* F1 was used as a non-target organism, and can degrade phenol but not 2,4-DCP. This organism has monooxygenase and dioxygenase genes, but the nucleotide sequences of these genes mismatch with the nucleotide sequence of the selected gene from reference organisms. The PCR products of amplified genes from reference microorganisms were cleaned with QIAquick PCR Purification Kit (Qiagen, Crawley, UK). In order to confirm the identity of the amplified products, the purified PCR products amplified from reference microorganisms were sequenced using the corresponding forward primer (3.2 pmol/µl), the ABI prism Big Dye Terminator Cycle Sequencing Ready reaction kit, and on an ABI Prism 377 DNA sequencer (Applied Biosystems, USA). Finally, the obtained nucleic acid sequences were compared against the Genebank database using BLAST algorithm to determine the closest matching sequence identity.

5.3.8 Quantitative real time PCR analysis for quantification of catabolic genes

Real time qPCR assays were developed and optimized in accordance with the MIQE (Minimum Information for Publication of Quantitative real time PCR experiments) guidelines, that provide the minimum information required to evaluate qPCR results in order to ensure their scientific integrity (Bustin *et al.*, 2009). The qPCR assay was performed on a BioRad CFX96 (Hercules, CA) with a C1000 thermal cycler iCycler and software version 3.0 (BioRad CFX Manager). qPCR reactions were prepared in 10 µL mixtures containing 2 µL of template DNA or cDNA (or 2 µl sterile molecular grade water for negative control, or 2 µl DNA from non-target organism to check non-specific amplification), 0.5 µL of 10 µM of each primer (Forward and Reverse), 2 µL sterile molecular grade water, with the 5 µL SsoFast EvaGreen Supermix for the CFX96 (Bio-Rad Labs Ltd, Hemel Hempstead, UK).

The following thermocycler program was used for amplifications in the qPCR assay: enzyme activation at 98 °C for 3 min, followed by 39 cycles of denaturation at 98 °C for 5 s and annealing at the primer specific annealing temperature for 5 s. The temperature gradient on the qPCR instrument was used to test different annealing temperatures (**Figure C 3a, Appendix C.7**). The qPCR products were subjected to melt curve analysis (**Figure C 3b, Appendix C.7**) at the end of the qPCR assay, where qPCR products were subjected to a gradual increase of temperature (0.2 ^{o}C temperature increments every 10 s) from 65 ^{o}C to 95 ^{o}C, and the corresponding rate of change in fluorescence signal intensities as a function of temperature was plotted. The annealing temperature with lowest C_t value and corresponding highest melt peak was considered the optimal temperature. Standard curves were constructed using the reference cDNA or nucleotide sequence of the target gene (details in the following section), and generated every time a qPCR analysis was performed, in parallel with the amplification of test samples. All samples were run in triplicate along with both positive (gene from reference microorganism) and negative (sterile molecular grade water and DNA from non-target microorganism) controls, using the optimized qPCR assay programs measured for each primer-set. The standard curves were generated by plotting the C_q values (i.e. the point at which the amplification curve crosses the threshold line / noise-band) versus the natural logarithm of the gene copy number (corresponding to the standard gene copies number; see below).

5.3.9 Catabolic gene standards for calibration of real-time PCR

PCR products of catabolic gene fragments from the reference microorganisms plasmid DNA were obtained with the respective designed primer-sets and purified using a QIAquick PCR purification kit (QIAGEN, Crawley,UK). The purified products were then quantified using the Quant-iT™ PicoGreen® dsDNA Assay Kit (Invitrogen, Life Technologies, Paisley, UK). The gene copy number of standard genes were estimated based on the PCR product size, assuming a molecular mass of 650 Da per base pair (**Appendix C.3**). Stock solutions of reference gene standards were prepared at a concentration of 10^9 gene copies/µL, and stored at -20°C until used. Calibration curves were prepared fresh for each qPCR run, by serially diluting (10-fold) the stock reference DNA standards to obtain standard solutions in the range of 10^8 - 10^1 target gene copies/µL.

5.3.10 Quantification of gene expression

Quantification of gene expression was conducted by synthesising complementary DNA (cDNA) from the total RNA using reverse transcriptase (RT), and using qPCR, with primer-

sets specifically targeting catabolic genes for the different enzymes (PH, C23D, *tfdB* and *tfdC*). The pure cultures of reference organisms were used to create the standards of reference cDNA.

5.3.10.1 cDNA standard preparation from pure culture

The total RNA was extracted from actively growing pure bacterial cultures (with either phenol or 2,4-DCP as the sole carbon substrate), as described in section **5.3.6**, and one-step RT PCR was performed using the QIAGEN OneStep RT-PCR kit (QIAGEN, Crawley,UK). This approach allows the inclusion of the reverse transcriptase step (for the cDNA generation) in the same tube as the PCR reaction, reducing the possibility of experimental errors. The RT-PCR reactions were performed in a total volume of 25 µL, containing 10 µM of each of the specific catabolic gene primer-sets (see **Table 5-2**), 1 pg - 2µg RNA template (or sterile molecular grade water, for the negative control), 10 mM of dNTPs, 5 µL of 5 x QIAGEN OneStep RT-PCR buffer and 1 ul of QIAGEN OneStep RT-PCR enzyme mix. The following thermocycler program was used for reverse transcriptase followed by PCR amplification: a reverse transcription at 50 ^{o}C for 30 minutes, initial PCR activation at 95 ^{o}C for 15 minutes, followed by 35 cycles of denaturation at 94 ^{o}C for 45 seconds, 1 minute at the primer specific annealing temperature (**Table 5-3**) and elongation at 72 ^{o}C for 1 minutes, with a final elongation step at 72 ^{o}C for 10 minutes. In order to confirm that the amplified cDNA were not from DNA contaminants present in the RNA template, separate reactions containing all the reagents and the original RNA samples as template were added in the thermocycler after the reverse transcriptase step as a control. The amplification of the correct size of RT-PCR products (i.e. cDNA) and their purity were confirmed by routinely performing electrophoresis, run at 100 V for approximately 1 hour, using 7 µL of product on 1% agarose gels containing 0.1 µL/mL concentration of Nancy 520 DNA stain (Sigma-Aldrich, UK), and in 1 x Tris-acetate-EDTA buffer. The RT-PCR products were then purified using a QIAquick PCR purification kit (QIAGEN, Crawley,UK), and quantified using the Qubit ssDNA kit (following the manufacturer's protocol) on a Qubit® 2.0 Fluorometer (Invitrogen, Life Technologies, Paisley, UK). The number cDNA copies of standard genes were estimated according to the RT- PCR product size, assuming a molecular mass of 650 Da per base pair (see details in **Appendix C.3**). The reference cDNA standards stock solution were prepared at a concentration of 10^9 gene copies/µL, and stored at -20°C until used. For calibration curve, cDNA standard ranging from 10^8 to 10^1 target gene copies/µL were prepared by making ten-

fold dilutions from the stock reference DNA standard. All RT-PCR products were stored at -20°C until used.

5.3.10.2 Complementary DNA from sludge samples

Synthetized cDNA was obtained from the total RNA extracted from activated sludge samples by reverse transcriptase, using the QuantiTect Reverse Transcription Kit (Qiagen, Crawley,UK). This involved the elimination of the genomic DNA from the total RNA sample, performed in a total volume of 14 µL, containing 2 µL of 7 x gDNA Wipeout mixture, template RNA (upto 1 µg) and RNase free water. The reaction tube was then incubated at 42°C for 2 min. This was immediately followed by the RT reaction step, where 14 µL of the obtained cDNA was mixed with 1 µL Reverse transcription master mix, 4 µl of 5 x Quantiscript RT buffer and 1 µL of RT primer mix. The following thermocycler conditions were used for the RT reaction: initial incubation at 42 ^{o}C for 15 minutes followed by inactivation of Quantiscript Reverse Transcriptase at 95 ^{o}C for 3 minutes. The final RT products were stored at -20^{o}C until further use.

5.3.11 Biodegradation assays for phenol and 2,4-DCP

The biodegradation potential of a series of activated sludge inocula sampled from different WWTPs was evaluated against a number of test chemicals using self-consistent experimental systems (batch system). The work developed according to the following steps:

1. Activated sludge was sampled from four wastewater treatment plants (WWTPs) and used as inocula for trial biodegradation assays to assess the biodegradation potential for phenol (5 mg/L) and 2,4-DCP (5 mg/L).
2. Total DNA samples isolated from the activated sludge inocula were tested for presence of catabolic genes involved in the biodegradation of the test chemicals. PCR-based assays were employed using specific primer-sets, designed in this work.
3. Activated sludge samples showing the presence of potential degraders were selected as inocula for further final biodegradation assays for phenol and 2,4-DCP. For biodegradation assays, phenol and 2,4-DCP were respectively spiked at the concentration of 50 mg/L and 5 mg/L. These selected concentrations are not inhibitory (Tyler and Finn, 1974; Abuhamed *et al.*, 2004).

5.3.12 Wastewater treatment plant sampling

Two litre grab samples of activated sludge were collected directly from the biological aeration basin of four selected WWTPs (**Table 5-4, Table 5-5 and Table 5-6**) and placed in

5 L bottles. Bottles were loosely capped during the transport and periodically shaken to maintain aeration. All the biodegradation assays were started within 2 hours of sampling.

5.3.13 Biodegradation Assays

Separate biodegradation studies were carried out for each of the test chemicals in batch reactor systems. The reactor consisted of 580 mL glass bottles capped with Mininert cap, inoculated with 200 mL activated sludge and spiked with either 50 mg/L of Phenol or 5 mg/L 2,4-DCP. Detailed description of the batch system is reported in **section 4.3.2, Chapter 4**. A control reactor for adsorption was prepared with activated sludge that was inactivated by autoclaving (121 °C for 20 min and 103 kPa) twice, 24 h apart (Helbling *et al.*, 2012), for each test chemical. In addition, an abiotic control was also prepared in sterile deionized water (DI) for each test chemical. All the reactors were kept at 22°C and continuously stirred (at 155 rpm) for the duration of the experiment. Traces of test chemicals (before spiking), total suspended solids (Clesceri *et al.*, 2005) and dissolved oxygen concentrations were measured at the beginning of the experiment. At regular time points, 2.5 mL of activated sludge inoculum was collected from the side port using a sterile syringe (Syringe Discardit II 2 mL; VWR, United Kingdom) and needle (VWR, United Kingdom). The sample at time zero was collected 2 minutes after chemical spiking, to assure complete mixing and dissolution of the chemicals. 0.75 mL of sample were centrifuged (8000 x g, 5 min) and filtered (13 mm diameter, 0.2 um PVDF syringe filter; VWR, United Kingdom) in order to remove solid material. 0.5 mL of the filtrate was mixed with an equivalent volume of methanol (Sigma-Aldrich, United Kingdom) and analysed to quantify the chemical concentration by HPLC (Shimadzu, UK). The detailed method and programs used for quantification of individual chemicals by HPLC is shown in **Appendix B, Table B 1**. 0.75 mL of sample was fixed in equivalent volume of 100 % ethanol, stored at -20 ^{o}C and later analysed in a flow-cytometer (Becton Dickinson, California) to quantify the total cell count according to the method mentioned below in **section 5.3.14**. The remaining 1 mL samples (0.5 mL for DNA extraction and 0.5 mL for RNA extraction) were immediately stored at -20 ^{o}C and -80 ^{o}C for DNA and RNA extraction respectively. The time zero sample for DNA and RNA extraction was collected prior to spiking the chemicals in the reactors.

A real time PCR method developed above (see section **5.3.8, 5.3.9 and 5.3.10** for details) was used to monitor the presence and expression of major catabolic genes associated with bacterial phenol and 2,4-DCP metabolism in the activated sludge. The *PH* and *C23D* genes, which respectively encode the *PH* and *C23D* enzymes were used as biological markers for

the meta-cleavage pathway in phenol biodegradation. Whereas for 2,4-DCP, *tfdB* and *tfdC* genes (encoding 2,4-DCP hydroxylase and chlorocatechol-1,2-dioxygenase enzymes respectively) were used. Catechol-1,2-dioxygenase is associated with the ortho-cleavage pathway in phenol biodegradation, which was not included in this study.

5.3.14 Quantification of total bacteria in activated sludge using flow cytometery

Total bacteria in activated sludge were quantified after disruption and dispersal of cells from flocs using a novel method (Brown *et al.*, 2018). Briefly, the stored samples were defrosted at room temperature and 900 µl of sample was mixed with 100 µL of pre-heated surfactant (5% Tween 80 and 10mM sodium pyrophosphate). Reagent blank was also prepared by adding 100 µL of surfactant to 900 ul of sterile DI water. Both samples and blanks were manually mixed for 10 s and then on a shaker at 200 rpm for 15 minutes, in the dark at room temperature. Finally, the samples and blanks were sonicated once in a water bath for 4 minutes, in cycles of 1 minute brusts and 1 minute rest.

The sample dilution was performed in TE-buffer (10 mM Tris-HCl 1 mM EDTA; pH 8.0) to achieve an event rate of between 200 and 800 bacteria s^{-1} and avoid coincidence (i.e. two or more bacteria being simultaneously counted within the sensing zone) (Brown *et al.*, 2018). To achieve this rate, five 1mL dilution series (1/200, 1/500, 1/750, 1/1000 and 1/1250) were prepared in triplicate per dilution. Diluted samples were stained with 10 ul of SYBR Green I (1:1000 dilution of commercial stock; Invitrogen, USA; λ_{ex} = 495 nm, λ_{em} = 525 nm; diluted in dimethyl sulfoxide (DMSO; Merck, Germany)) for 10 min in the dark at 60 ^{o}C.

The total bacteria were quantified using FACScan flow cytometer (Becton Dickinson, California) equipped with a 15-mW 488-nm air-cooled argon-ion laser and a standard filter setup. Readings were collected in logarithmic mode and analysed with Flowing Software 2.0. Electronic gating was used to separate selected signals (prokaryotic cells) from background (inorganic and organic particles)(Berney *et al.*, 2007; Hammes *et al.*, 2008)

5.4 Results

5.4.1 Specificity of the primer-sets

The specificity of the three newly designed primer-sets and one from the literature were examined by evaluating the PCR products amplified from the plasmid DNA of two reference strains mentioned in **Table 5-1**, of a non- target bacterial culture DNA and from total DNA of the activated sludge, using the primer-set for each catabolic gene at its optimized annealing temperature. The primer-sets were found to be specific to the target genes; visualization of PCR and RT-PCR (cDNA) on agarose gel showed a single band of the expected size without

(**Figure C 2, Section 2, Appendix C.6**) amplification of any fragments of DNA from the non-target bacterium culture (i.e. *Pseudomonas putida* F1). Furthermore, sequencing of the amplified DNA fragments from reference microorganisms confirmed that there was a match with that of the target gene (**Figure C 1**, **Appendix C.5).**

5.4.2 Quantitative real-time PCR (qPCR) optimization

The optimum annealing temperature for the four primer-sets was determined separately for qPCR assays by testing temperatures above and below the optimum annealing temperature determined previously for the general PCR reaction, and using known standard concentrations (10^5 gene copies/µL) as templates (see section **5.3.8, 5.3.9 and 5.3.10)**. The melt curve analysis performed for all reference catabolic genes showed a single characteristic peak for respective the amplicon, at the reaction conducted at optimized annealing primer temperature (**Appendix C.7, Figure C 9b**). The qPCR assay products were further analysed on an agarose gel, and confirmed a single amplicon was generated which correspond to the observed single peak in the melt curve (see **Figure C 3b, Appendix C.7**).

Calibration curves were produced for eight reactions [four total DNA (i.e. four genes) and their respective cDNA qPCR assays] in the range of 10^8 to 10^1 gene copies μL^{-1} of qPCR standards for each catabolic gene. Standard curves (**Table 5-3**) indicated that amplification efficiency was high (89.68 - 110.18 %), which in practice should be between 90 – 110 % (Dorak, 2006). These results confirm that the qPCR reactions progressed well with good sensitivity. The standard curves were linear (R^2 >0.99) over 8 orders of magnitude from 10^8 to 10^1 gene copies μL^{-1} for all standards.

Table 5-3 Summary of an optimum annealing temperature and efficiency of the primers used in this study for qPCR and RT-qPCR assays

Primer-sets	T_a (ºC)	Slope (m)	% Efficiency ($100*[10^{(-1/slope)} – 1]$)	R^2
Phenol hydroxylase (*PH*)	59.4	-3.57	90.568	0.998
PH cDNA	57.6	-3.17	106.718	0.987
Catechol-2,3-dioxygenase (*C23D*)	50	-3.51	92.831	1.000
C23D cDNA	51.9	-3.56	90.813	0.998
2,4-DCP hydroxylase (*tfdB*)	54.5	-3.10	110.175	0.999
tfdB cDNA	57.6	-3.35	98.84	0997
chlorocatechol-1,2-dioxygenase (tfdC)	60.7	-3.60	89.675	0.998
tfdC cDNA	59.2	-3.44	95.435	0.999

T_a; Primer annealing temperature

5.4.3 Biodegradation assay

5.4.3.1 Trial tests

83.2 (s.d.+/- 5.16) % of spiked phenol was degraded within 1.5 hours by activated sludge inocula from all WWTPs (**Figure 5-2**), while efficient 2,4-DCP degradation was only observed when using an activated sludge inoculum from one of the four WWTPs (see **Figure C 4, Appendix C.8**), where 96.42 % of spiked 2,4-DCP degraded within 162 hours. Furthermore, PCR confirmed the presence of catabolic genes associated with phenol degradation in total DNA extracted from all activated sludge samples (**Section 4**, **Appendix C.6**). Whereas, the catabolic genes involved in 2,4-DCP degradation were only detected in the plant supporting 2,4-DCP degradation (**Section 6 [Figure C 2]**, **Appendix C.6**). Therefore, samples from WWTP4 were used to further study 2,4-DCP biodegradation in activated sludge.

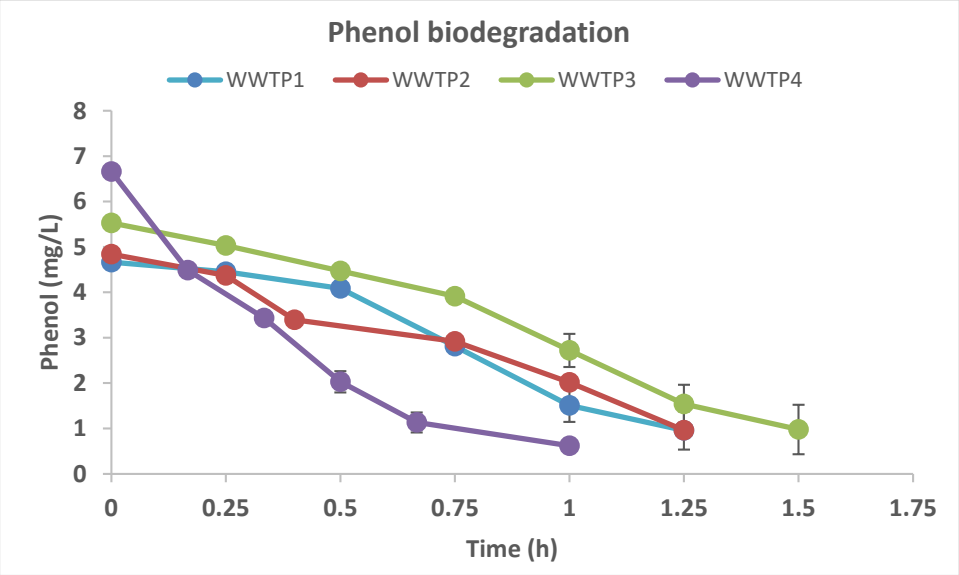


Figure 5-2 Biodegradation of phenol in activated sludge sampled from different WWTPs. Error bars represent the standard deviations for duplicate experiments and might not be visible in some cases

Table 5-4 List of WWTPs, description, properties, abundance of initial phenol catabolic genes and first-order phenol biodegradation rates in trial experiments

Trial experiment for phenol biodegradation							
WWTP Site	**Influent source**	**Total Suspended Solid (gTSS/L)**	**Dissolved Oxygen (mg/L)**	**pH**	**Total catabolic gene copies/mL AS[b]**	**1st order rate (h^{-1})**	**Sampling Date**
WWTP1	Domestic	3.56 (0.523)[a]	3.57	6.37	1.44×10^6	1.57 (0.056) [a]	02/11/2016
WWTP2	Domestic	4.7 (0.1) [a]	224	7.17	1.29×10^6	1.19 (0.002) [a]	02/12/2016
WWTP3	Domestic + Industrial	2.32 (0.102) [a]	2.95	6.75	1.34×10^6	1.22 (0.336) [a]	31/10/2016
WWTP4	Industrial	4.94 (0.467) [a]	3.41	6.45	7.09×10^7	2.47 (0.017) [a]	19/12/2016

( )[a] : standard deviation for the duplicate experiments are shown in brackets. [b] AS: Activated sludge

Table 5-5 List of WWTPs, description, properties, abundance of initial 2,4-DCP catabolic genes and first-order 2,4-DCP biodegradation rates in trial experiments

Trial experiment for 2,4-DCP biodegradation							
WWTP Site	**Influent source**	**Total Suspended Solid (gTSS/L)**	**Dissolved Oxygen (mg/L)**	**pH**	**Total catabolic gene copies/mL A.S**	**1st order rate (h^{-1})**	**Sampling Date**
WWTP1	Domestic	3.355 (0.277) [a]	2.47	6.75	not detected	0.0033 (0.000141)	06/12/2016
WWTP2	Domestic	4.7 (0.1) [a]	2.24	7.17	not detected	0.00205 (0.0001)	02/12/2016
WWTP3	Domestic + Industrial	2.113 (0.093) [a]	2.85	7.14	not detected	0.00205 (0.00354)	22/12/2017
WWTP4	Industrial	6.726 (0.219) [a]	5.48	6.69	6.45×10^5	0.00930 (0.0005) [a]	19/12/2016

()[a] : standard deviation for the duplicate experiments

Table 5-6 List of WWTPs, description, properties, abundance of initial phenol and 2,4-DCP catabolic genes and first order biodegradation rate in final experiments

Final Experiment							
Chemical	**WWTP Site**	**Influent source**	**Total Suspended Solid (gTSS/L)**	**Dissolved Oxygen (mg/L)**	**pH**	**Total catabolic gene copies/mL A.S**	**1st order rate (h^{-1})**	**Sampling Date**
Phenol	WWTP1	Domestic	3.744 (0.245) [a]	2.75	6.55	2.76×10^6	0.4 (0.013) [a]	22/12/2017
2,4-DCP	WWTP4	Industrial	4.755 (0.673) [a]	3.54	6.6	7.34×10^5	0.0189 (0.003) [a]	09/01/2017

()[a] : standard deviation for the triplicate experiments

5.4.3.2 Chemicals biodegradation and catabolic genes abundance

Phenol (50 mg/L) and 2,4-DCP (5 mg/L) were effectively removed by the microbial community present in the activated sludge from WWTP 1 and WWTP 4, respectively, with degradation occurring within 15 hours for phenol (**Figure 5-3**) and approximately 7 days for 2,4-DCP (**Figure 5-4**). No major phenol losses were recorded due to adsorption or abiotic processes, as shown in **Figure 5-3.** Approximately 75% of the spiked 2,4-DCP was measured at time zero in the liquid phase of the reactors containing activated sludge, suggesting an instant adsorption of 2,4-DCP on the sludge surface. This was confirmed by the initial measurements in the abiotic control reactors, which showed a 100% recovery at time zero.

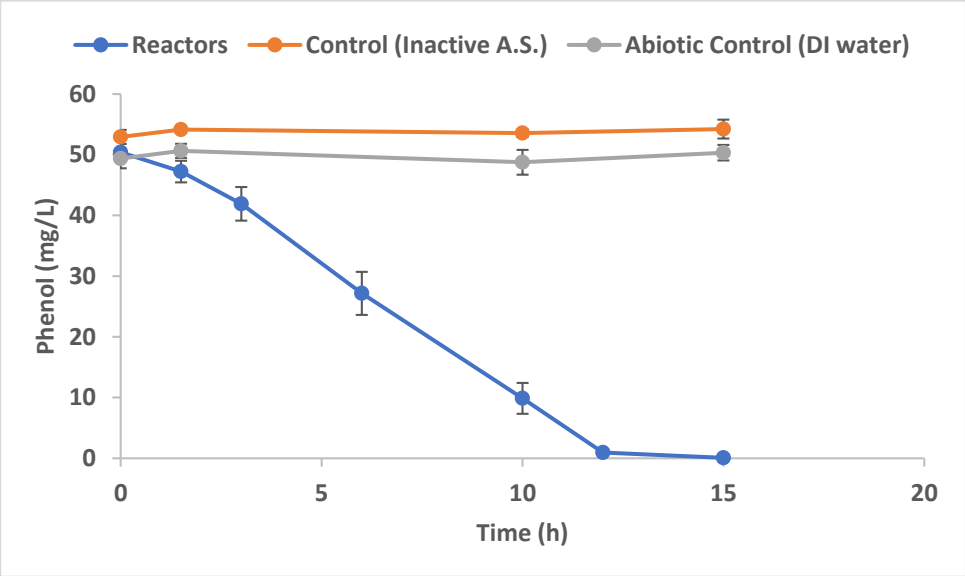


Figure 5-3 Biodegradation of phenol in activated sludge. Experiments were conducted in triplicate with two controls to check adsorption of phenol on solids and abiotic degradation of phenol. For control experiments, a single reactor was run, while sampling was perfored in duplicate at each time point. Error bars represent the standard deviations for experiments and might not be visible in some cases.

Analysis of total DNA isolated from phenol- and 2,4-DCP-amended activated sludge (final biodegradation assays) by real time PCR revealed the presence of major catabolic enzymes involved in phenol (PH and C23D) and 2,4-DCP (*tfdB* and *tfdC*) degradation at all the time points. The key catabolic genes concentrations increased over the duration of experiment for both chemicals (**Figure 5-5** and **Figure 5-6**) where the final catabolic gene concentrations (after 15 and 160 hours) were significantly more abundant than the initial ones (P-value < 0.05, 2-sampe t-test). In addition, chemical concentration decreased with increased gene copy numbers for all studied genes.

In contrast, the abundance of total bacteria remained similar or slightly lower than the initial abundance over the duration of experiment for both chemicals (**Figure 5-5** and **Figure 5-6**). All the catabolic genes were detected at time zero, indicating that microorganisms possessing

the metabolic capability to degrade the phenol, 2,4-DCP and their metabolites were present in the respective activated sludge inocula before the application of the pollutant.

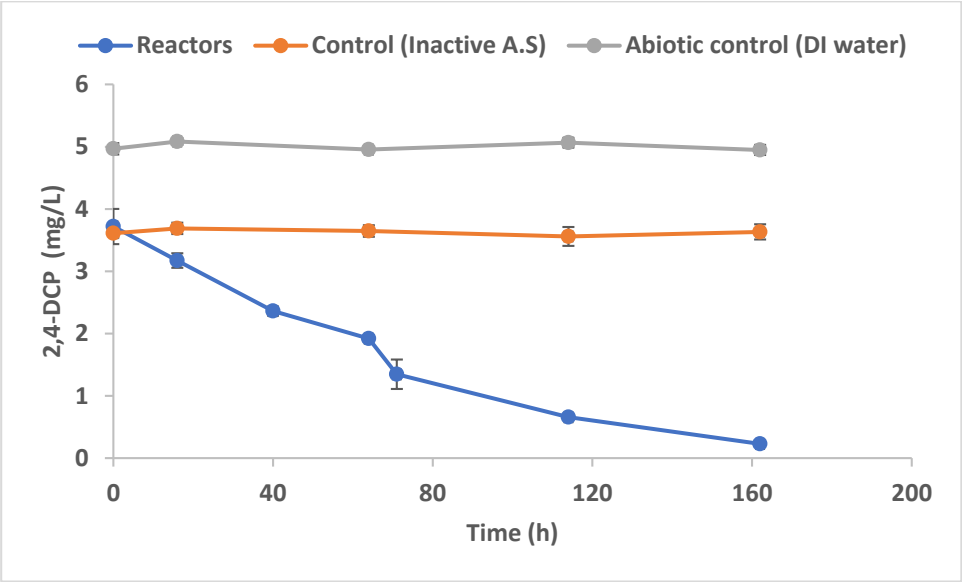


Figure 5-4 Biodegradation of 2,4-DCP in activated sludge. Experiments were conducted in triplicate with two controls to check adsorption of phenol on solids and abiotic degradation of phenol. For control experiments, a single reactor was used, while sampling was performed in duplicate at each time point. Error bars represent the standard deviations for experiments and might not be visible in some cases.

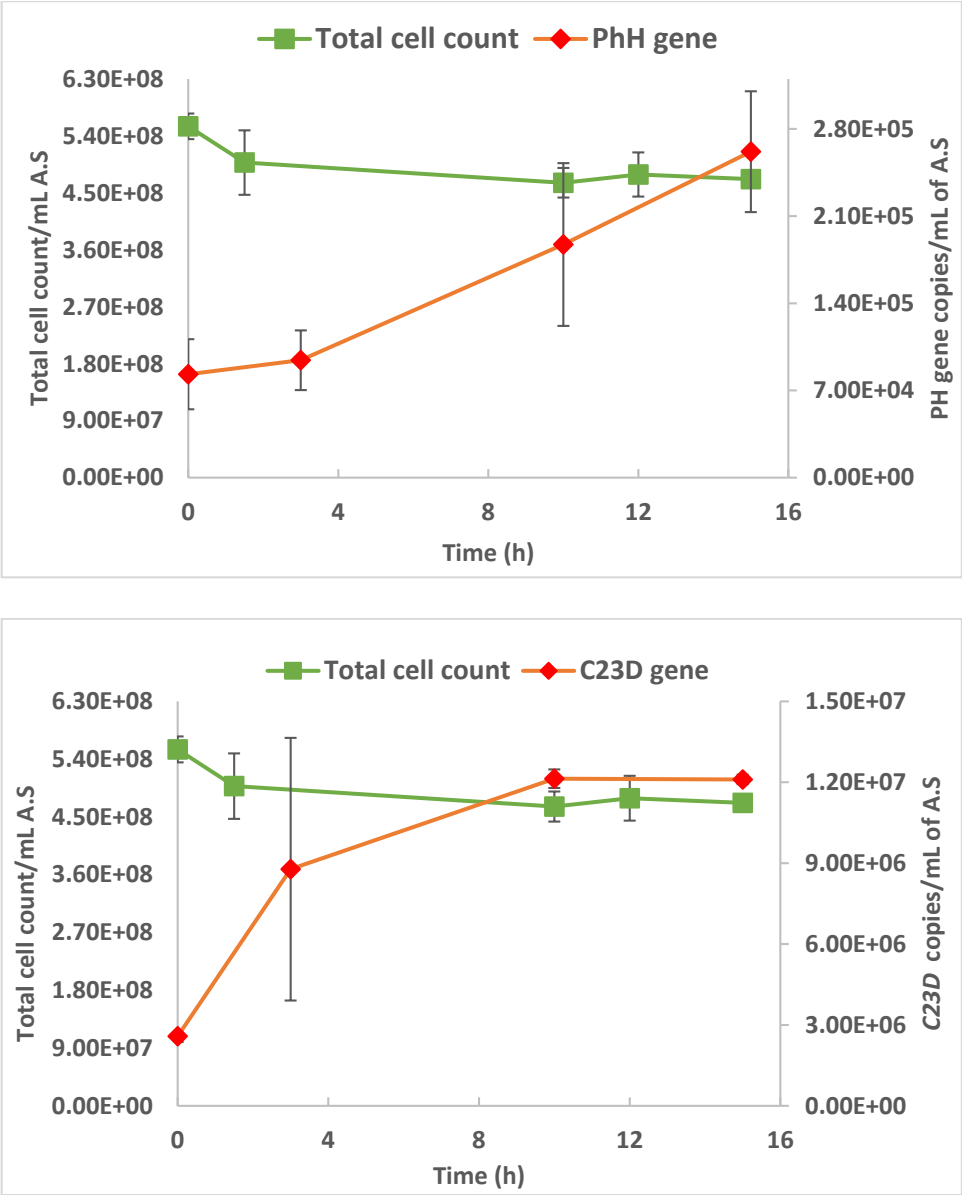


Figure 5-5 Total bacterial counts and phenol catabolic gene (C23D and PH) copies over time for the phenol biodegradation experiment conducted with activated sludge inocula sampled from WWTP1. Data points are an average of biological triplicate samples. Error bars are ± 1 s.d. of biological triplicate measurements.

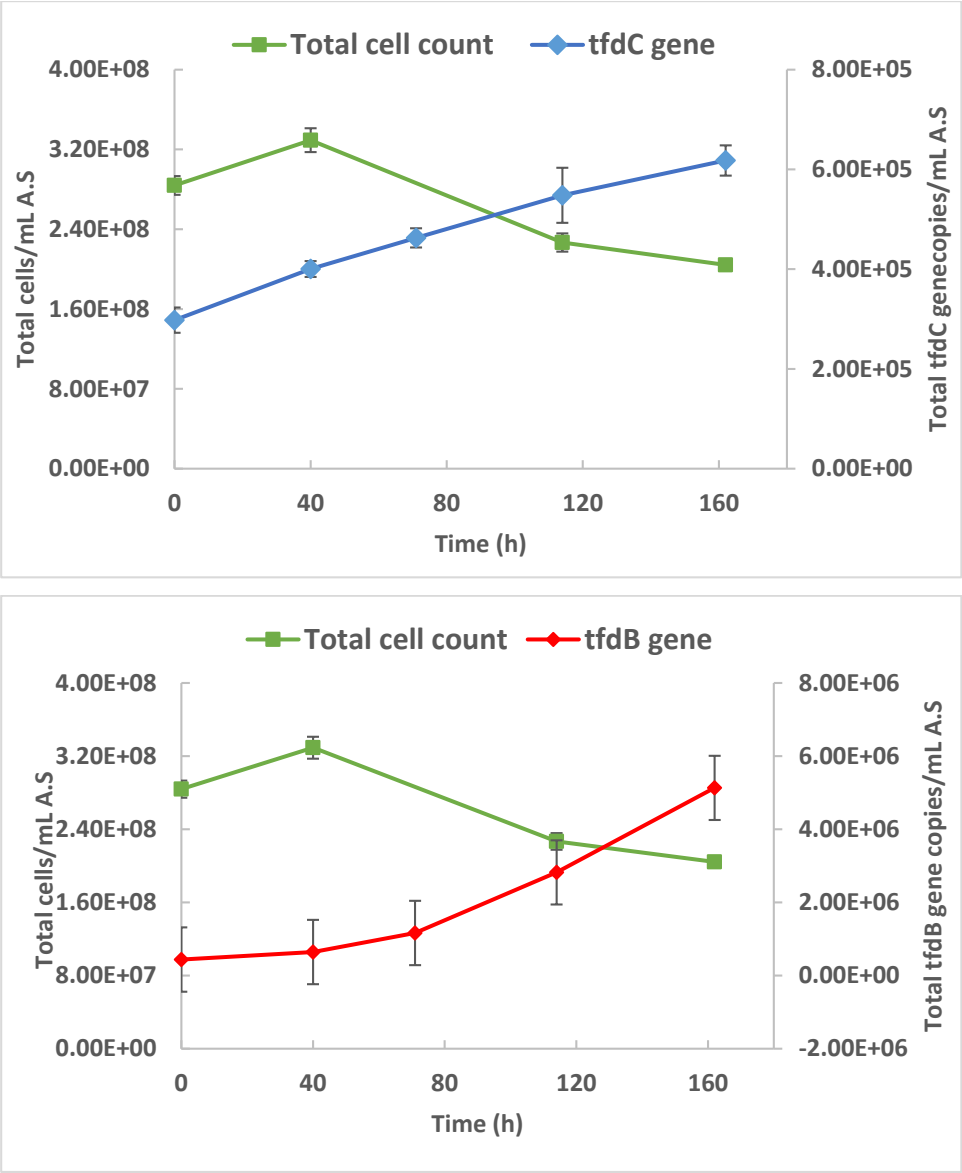


Figure 5-6 Total bacterial counts and 2,4-DCP catabolic gene (*tfdB and tfdC*) copies over time for the 2,4-DCP biodegradation experiments conducted with activated sludge inocula sampled from WWTP4. Data points are an average of biological triplicate samples. Error bars are ± 1 s.d. of biological triplicate measurements.

5.4.3.3 Catabolic gene expression during phenol and 2,4-DCP degradation

Phenol and 2,4-DCP were not detected in the aqueous phase of the activated sludge samples used for the biodegradation experiments. It has to be noted that, the limit of detection, which indicates the lowest quantity or concentration of a component that can be reliably detected with a given analytical method, for phenol and 2,4-DCP were 0.01 mg/L and 0.002 mg/L, respectively. The expression profiles of PH, C23D, *tfdB* and *tfdC* genes from RT-qPCR experiments are shown in **Figure 5-7** and **Figure 5-8**. After spiking phenol in the reactor, PH and C23D genes were induced, and resulted in expression of these genes and phenol degradation was observed. The mRNA for phenol hydroxylase gene was not detected at time zero in the extracted total RNA from activated sludge, while mRNA for C23D was detected, potentially suggesting that the presence of phenol, or other, metabolites (i.e. catechol) were present in the activated

sludge prior to spiking the phenol in the reactor. Alternatively, C23D transcripts were produced constitutively in the activated sludge used in this study. It is noteworthy that the expression of the C23D gene was significantly abundant (2-sample t-test, $P< 0.05$) compared to the expression of PH gene. However, the expression pattern of these genes was similar, showing a steady increase over the duration of experiment. The PH and C23D gene expression was significantly lower at 3 h, but increased by 76- and 10- fold, respectively after 10 h. In the case of the 2,4-DCP biodegradation experiment, mRNA of *tfdB* and *tfdC* genes were detected at time zero, theoretically suggesting presence of 2,4-DCP and its metabolite in the activated sludge prior to spiking the chemical. However, 2,4-DCP was not detected in the initial activated sludge (Limit of Detection for 2,4-DCP was 0.002 mg/L) and this suggests that *tfdB and tfdC* transcripts were produced constitutively at a high level in the activated sludge used in this study. Both genes were expressed significantly after spiking 2,4-DCP in the reactor, confirming that these genes were involved in the 2,4-DCP biodegradation pathway. These genes represent the major catabolic genes in the 2,4-DCP degradation pathway, and are involved in the aromatic ring cleavage reaction steps. The expression of *tfdC* and *tfdB* genes was low at time 0, but significantly increased 11.5 - fold after 71 h before coming down to about 9.91 – fold after 162 h. The gene expression data showed good linear correlation between the expression of these genes and the amount of substrate removed during the experiments (PH and C23D gene expression; $r^2 = 0.98$ and *tfdB* and *tfdC* gene expression; $r^2 = 0.78$).

Table 5-7 Ratio of total catabolic gene transcripts and total catabolic genes over time during the biodegradation of phenol and 2,4-DCP in activated sludge sampled from WWTP1 and WWTP4, respectively

Chemical	Time (h)	Ratio of catabolic gene transcripts/ catabolic genes
2,4-DCP	0	13.38
	40	42.52
	71	71.52
	162	16.93
Phenol	0	0.01
	3	0.47
	10	5.56

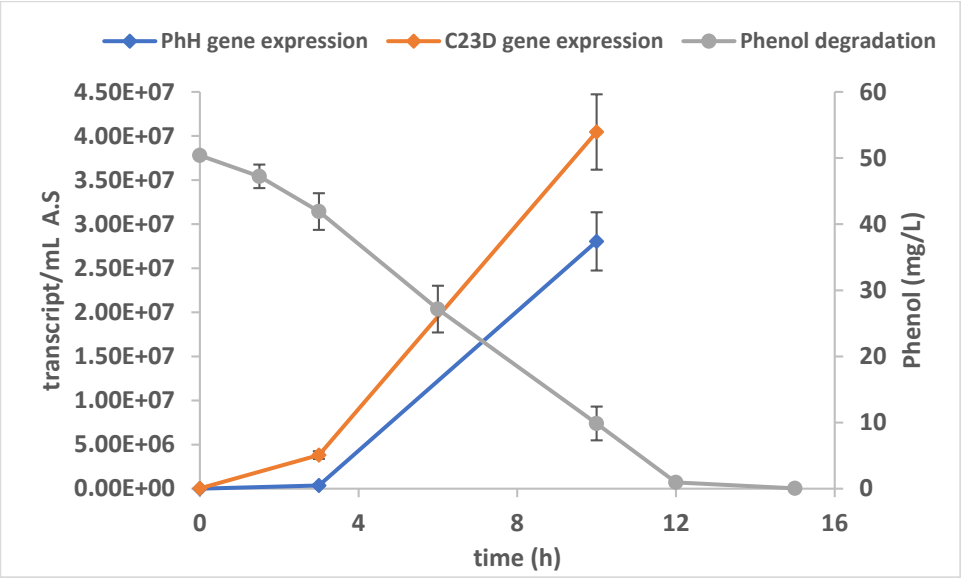


Figure 5-7 Phenol removal and expression profile of PH and C23D genes over time during the biodegradation of phenol using an activated sludge inoculum sampled from WWTP1. Data points are an average of biological triplicate samples. Error bars are ± 1 s.d. of biological triplicate measurements.

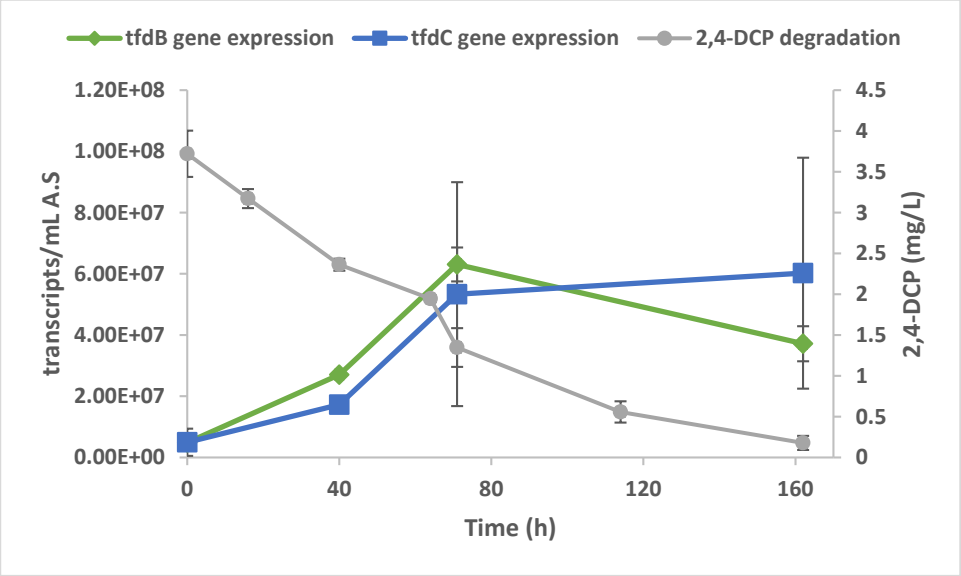


Figure 5-8 2,4-DCP removal and expression profile of *tfdB* and *tfdC* genes over time during the biodegradation of 2,4-DCP using an inoculum of activated sludge sampled from WWTP4. Data points are an average of biological triplicate samples. Error bars are ± 1 s.d. of biological triplicate measurements.

5.5 Discussion

In this study, the major catabolic genes involved in the biodegradation of some selected aromatic chemicals were identified by using a pathway prediction system. The identified genes were used as biological markers to assess the biodegradation potential of an environmental system for a given chemical. A pair of primer-sets able to detect and quantify each of the identified catabolic genes in the environmental system were designed and/or selected from the literature. Molecular microbial assays (qPCR and RT-qPCR) were developed and optimized for each primer-set to assess and quantify the activity of potential

degraders in an environmental system during the biodegradation process. Combining the approach used in this study with biodegradation kinetic models might enable better prediction of the biodegradation rate of organic pollutants in the environment. However, it should be noted that the functional gene approach used in this study has limitation, mostly primer coverage. The coverage of the primer-sets in terms of capturing catabolic genes performing specific functions in the environment is, however, strongly affected by the inclusion or exclusion of nucleotide sequences of commonly involved alleles in the degradation pathways in the library used for primer design. The primers used in this study were designed by aligning the homologous sequences and thus only able to detect and quantify the degraders closely related to reference microorganisms, which was used for library generation for primer design. However, there are numerous other degraders in the environment possessing the monooxygenase and dioxygenase genes, whose sequences are not homologous to the catabolic gene sequences of reference microorganisms, and still have an ability to degrade chemical. For example, *Pseudomonas putida* F1 has a group of these aforementioned genes, which are not homologous to phenol hydroxylase and catechol-2,3-dioxygenase genes (**Figure 5-1**), but still able to degrade phenol efficiently (Chapter 4). Therefore, it is most likely that primers used in this study under-quantified the potential putative test chemical degraders than that actually present in the environment due to low coverage.

5.5.1 Primer design

The primer-sets were successfully applied to evaluate and quantify the response of specific degraders in activated sludge during test chemical degradation. Degenerate primer-sets have been show to target wide range of genes with same function in total DNA extracted from environmental system (Cébron *et al.*, 2008; Meynet *et al.*, 2015), a finding that has been confirmed by these results (**Appendix C.5 and Appendix C.6**). Designed and literature-adapted primer-sets showed good specificity to the target genes in the environmental total DNA and reference organisms. The sequencing results indicated that the designed primer-sets were not biased towards any one specific allele involved in the degradation pathways, it rather amplified the commonly involved alleles in the degradation pathway of test chemicals. However, the presence of other similar or unknown genes that have the potential to carry out the same function as the genes targeted by the primer-sets used in this study cannot be discounted; the primer-sets used here might not have been able to identify them in environmental inocula. On the other hand, non-specific amplification, which is common in PCR amplification with environmental DNA was not observed. This was indicated by the

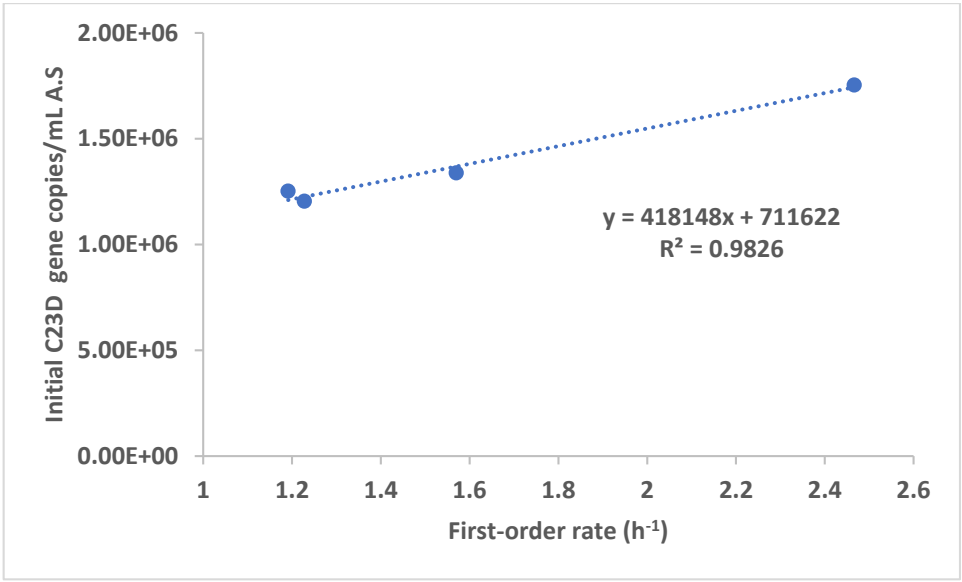
appearance of melt curve peak for both positive control and sample DNA at same position during qPCR analysis (**Figure C 9, C 10 and C 11, Appendix C.12**).

5.5.2 qPCR Assay

The real-time qPCR assays developed in the present study allowed quantification of catabolic genes directly from the environmental samples. Calibration standards ($10^8 – 10^1$) for all studied genes were within the linear range and suggests that the detection limit for all the studied genes were low. The quantification accuracy of qPCR can be greatly influenced by number of factors, that includes DNA extraction and recovery efficiency (affected by cell wall characteristics of organisms), the presence of PCR inhibitors in genomic DNA extract (humic acid and organic matters) and differences in the amplification efficiency for the standard and the sample target template (Van Doorn *et al.*, 2009; Brankatschk *et al.*, 2012). To minimize the effect of PCR inhibitors and improve the qPCR efficiency, the genomic DNA samples were diluted to the concentration showing minimum inhibition during qPCR amplification (result not shown). However, the presence of PCR inhibitors that may have influenced the qPCR quantification efficiency cannot be ruled out. While, one point calibration, which accounts for template related variability of efficiency by correcting for differences in efficiency between sample target template and standard, might be used instead of standard curve method for absolute quantification (Brankatschk *et al.*, 2012).

5.5.3 Trial biodegradation assays

The observed degradation of phenol and detection of major catabolic genes involved in its degradation pathway in activated sludge sampled from all the WWTPs suggest that phenol degraders may be ubiquitous in WWTPs. However, the rate of phenol degradation differed across plants. Such differences may be due to differences in different biotic or abiotic parameters of the activated sludge (e.g. dissolved oxygen (DO) concentration, the initial degrader numbers, total suspended solids - TSS, pH, types of degraders) (**Table 5-4**, **Table 5-5** and **Table 5-6**). Interestingly, a linear correlation [p-value < 0.05; $r^2 = 0.9819$ (PH gene) and 0.9826 (C23D gene)] was observed between the initial catabolic gene copy numbers and phenol biodegradation rates across different WWTPs (**Figure 5-9**). (Hickman and Novak, 1989) demonstrated a direct (positive) relationship between biodegradation rates of phenol in surface soil and bacterial densities measured by aerobic plate counting techniques .



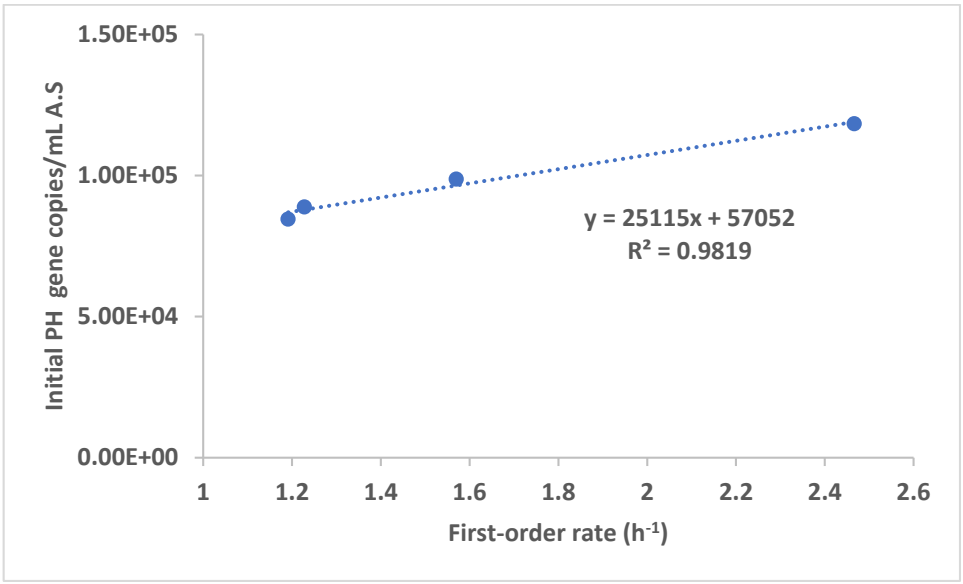


Figure 5-9 Univariate regression analysis between first order biodegradation rate of phenol and initially observed PH (bottom) and C23D (top) catabolic genes for phenol biodegradation in activated sludge inocula from different WWTPs.

2,4-DCP degradation was observed when using an activated sludge inoculum from only one WWTP where *tfdB* and *tfdC* degrading genes were detected. This result indicates that 2,4-DCP degraders are not ubiquitous like phenol degraders, and the presence of specific catabolic genes and their expression is an important factor for efficient chemical biodegradation. This finding was further supported by an experiment that demonstrated that 2,4-DCP degradation was only observed when known degraders were spiked into a reactor that contained an inoculum of activated sludge (WWTP1) devoid of such catabolic genes, where 2,4-DCP degradation was previously not observed (**Figure C 4b, Appendix C.8**).

5.5.4 Gene abundance and expression

The catabolic genes used in this study were selected not only because they indicate the pathway for degradation of key phenol and 2,4-DCP metabolites (phenol ; catechol and 2-

hydroxymuconic semi-aldehyde and 2,4-DCP: 3,5-dichlorocatechol and 2,4-dichloro-cis,cis-muconate) (Gao *et al.*, 2010), but also because they are often the rate limiting step in aromatic chemical biodegradation (Pitter and Chudoba, 1990); hydroxylation of the aromatic ring with subsequent ring cleavage. Thus, these genes may be useful in monitoring the in-situ sequential degradation of phenol and 2,4-DCP.

Phenol hydroxylase genes, which initiate the phenol degradation reaction (Gao *et al.*, 2010), were present in significantly lower amounts than the catechol-2,3-dioxygenage gene, which is involved in the second reaction step (i.e. ring cleavage reaction)(Gao *et al.*, 2010). Catechol (or its moieties), which is the intermediate of first step reaction, is also the intermediate biodegradation product of other chemicals like benzene, toluene and cresols (Gao *et al.*, 2010), and is utilized by other aromatic carbon-degrading microorganisms. This implies the primer-set for catechol-2,3-dioxygenase gene is not specific for phenol degradation, but also targets microorganisms that can metabolize other chemicals. However, it also cannot be assumed that most microorganisms are able to grow on only one aromatic chemical. For example, *Pseudomonas putida* F1 (Reardon *et al.*, 2000) and *Pseudomonas sp*. CF600 *(Powlowski and Shingler, 1994)* are perhaps the best characterized microorganisms that are able to use other aromatics (phenol, cresols, BTEX) as sole carbon and energy sources. In organisms like these, many of their enzymes can be induced in a non-specific in nature. As a result, many of their metabolic pathways contain a high degree of convergence (Hutchinson and Robinson, 1988), which therefore allows them the opportunity to efficiently utilize a wide range of chemicals without an excess of redundant genetic coding for enzyme induction. Due to this, the use of primer-sets for phenol catabolic genes in quantification of phenol degraders in an environmental system might not always quantify the true number of phenol degraders.

Likewise, dichlorophenol hydroxylase (*tfdB*) and chlorocatechol-1,2-dioxygenase (*tfdC*) are the first and second enzymes respectively involved in steps for aerobic 2,4-DCP metabolism. Many studies have shown these genes to facilitate the metabolism of halogenated aromatic chemicals (Chaudhry and Chapalamadugu, 1991; Cavalca *et al.*, 1999; Lillis *et al.*, 2010). The quantification of *tfdB* and *tfdC* genes from different WWTPs indicate that, unlike phenol degraders, 2,4-DCP degrading microorganisms are not ubiquitous in nature. In this study, 2,4-DCP degraders were identified in activated sludge sampled from a WWTP treating industrial wastewater and were absent in WWTPs treating domestic sewage. It should also be noted that, the wastewater coming from industries served by this WWTP mostly constitute solvents, xenobiotics, herbicides, and degreasers among other industrial chemicals (Sánchez-Avila *et al.*, 2009). This can ultimately influence the shape and composition of activated sludge

microbial communities and might promote the efficient removal of non-readily biodegradable chemicals like 2,4-DCP.

The observed increase in the catabolic gene abundances and their expression over the duration of the experiment indicates that the degradation of these chemicals in the reactor is probably controlled by both enzyme regulation and enrichment of competent cells. In several other studies (Alfreider *et al.*, 2003; Lillis *et al.*, 2010), active degraders involved in aromatic chemical degradation actively expressed their catabolic genes upon exposure to aromatic chemicals, with or without enrichment of competent cells. However, the enrichment or growth of organisms upon exposure to chemicals depends on the initial chemical concentration (Kovárová-Kovar and Egli, 1998).

It has to be noted that only a small fraction of the microbial cells present in the activated sludge microbial communities have the potential to degrade certain pollutants (Kovárová-Kovar and Egli, 1998) and their enrichment during the exposure to pollutant might not significantly increase the final total number of bacterial cells in the system. Our total cell count results for both test chemicals are in agreement with this. In the case of 2,4-DCP degradation, the final total cell count was significantly lower than the initial cell count. This might be because the biodegradable carbon and available dissolved oxygen in the activated sludge became limited and ran out as the experiment proceeded, eventually leading to cell death and lysis, thereby releasing the residual internal molecules, which then became available as a potential food source for other bacteria.

5.6 Conclusion

Combining traditional methods for biodegradation studies with molecular techniques used in this study will improve our understanding of the composition and physiology of metabolically active members of the microbial community in an environment. In addition, it will also help to access the biodegradation potential of an environment and may also help to evaluate the overall kinetics of the chemical biodegradation. Overall these results indicate that the pollutants used here were eliminated biologically and this was the result of both the enrichment of key degraders and expression of relevant genes in potential degraders present in indigenous populations. Even though phylogenetically diverse groups of bacteria are present in activated sludge, the catabolic genes involved in the biodegradation of some chemicals (eg., 2,4-DCP) are not always present. The presence or absence of potential chemical degraders possessing the catabolic ability to degrade particular chemicals will ultimately determine the degradation potential an environment for such a chemical.

Chapter 6

Can initial competent degrader abundance and chemical concentration be used to predict biodegradation rate of aromatic chemicals in activated sludge?

Chapter 6. Can initial competent degrader abundance and chemical concentration be used to predict biodegradation rate of aromatic chemicals in activated sludge?

6.1 Abstract

A factorial design approach was used to investigate the effect of putative degrader numbers in inocula and initial chemical concentration on the biodegradation rate of phenol and 2,4-dichlorophenol (2,4-DCP) in activated sludge using laboratory batch experiments. An empirical regression model predicting the first order biodegradation rate of phenol and 2,4-dichlorophenol was built for both chemicals based on the number of specific degraders and concentration of the chemical at the start of the experiment. The model could explain much of the variation of the biodegradation rate with R^2 values of 97.9% and 97.8 % respectively for phenol and 2,4-DCP. The initial specific degrader numbers (X_0) had a positive effect on chemical biodegradation rate, whereas the initial chemical concentration (C_0), and the interaction between X_0 and C_0, had a negative effect on biodegradation rate. For both chemicals, strong and significant linear correlations [Pearson correlation coefficient (r) > 0.9 and p-value < 0.05] were observed between biodegradation rate and the X_0/C_0 ratio, while no strong correlation was detected between biodegradation rate and initial chemical concentration and specific degrader numbers alone. The results derived from this study indicate that the biodegradation rate of the chemicals in natural and engineered ecosystems is a function of both intrinsic properties of the chemical (i.e. chemical concentration) and extrinsic characteristics of the environment (e.g. abundance of specific degraders in the inoculum).

6.2 Introduction

Biodegradation is an important but poorly understood fate process that affects persistence assessments for regulatory purposes and could affect the design and operation of biological wastewater treatment systems for achieving effective removal of chemical pollutants (Pavan and Worth, 2008; Goodhead *et al.*, 2014; Martin *et al.*, 2017b). In particular, the rate of biodegradation is difficult to predict as it is influenced by many different factors, among which the concentration of chemicals, physico-chemical conditions, and the concentration of putative chemical degraders have been proven to be the most important (Thouand *et al.*, 1996; Kowalczyk *et al.*, 2015b). The development of models for the prediction of biodegradation rates of chemicals that incorporate such factors is therefore essential.

Theoretically, the kinetics of chemical catabolism (i.e. chemical biodegradation rate) of the pure culture and mixed culture can be described by various kinetic models, where concentration of the introduced substrate (i.e. chemical) and abundance of active degraders

are the two key variables (Simkins and Alexander, 1984; Simkins and Alexander, 1985; Okpokwasili and Nweke, 2006). First order, zero-order, logistic, logarithmic, Monod (no growth) and Monod (with growth) kinetics are different types of proposed kinetic models that use such variables (**Table 6-1**).

All the models mentioned in **Table 6-1** are considered to be the special or simplified forms of the integrated Monod equation [**Equation 1 (section 4), Appendix D**], which was proposed by Alexander and Simkins (1984), at extreme ratios of initial chemical concentration to half - saturation constant or at the extreme ratios of initial active degrader abundance to initial chemical concentration (Simkins and Alexander, 1984). In this study, chemical degradation rates were estimated by fitting biodegradation into different kinetic equations, and the one with the best R^2 value was used for model development.

Table 6-1 Biodegradation model for mineralization kinetics derived from integrated Monod Equations with the variable of substrate concentration and inocula concentration (Reproduced from:(Simkins and Alexander, 1984))

Biodegradation models	**Integral form**
Zero order Necessary condition	$C = C_0 - k_1 * t$ ($X_0 >> C_0$ and $C_0 >> K_s$)
Monod, no growth Necessary condition	$K_s \ln(C/C_0) + C - C_0 = -k_1 * t$ ($X_0 >> C_0$)
First order Necessary condition	$C = C_0 \exp(-k_1 * t)$ ($X_0 >> C_0$ and $C << K_s$)
Logarithmic Necessary condition	$C = C_0 + X_0 [1 - \exp(\mu_{max} * t)]$ $C_0 >> K_s$
Logistic Necessary condition	$C = (C_0 + X_0)/(1 + (X_0/C_0) \exp[k_1(C_0 + X_0)t])$ ($C_0 << K_s$)
Monod with growth Necessary condition	$K_s * \ln(C/C_0) = (C_0 + X_0 + K_s)\ln(X/X_0) - (C_0 + X_0)\ \mu_{max} * t$ none

C: chemical concentration, C_0: initial chemical concentration, K_s: half-saturation constant for growth, X: inocula concentration, X_0: initial inocula concentration, μ_{max}: maximum specific growth, t: time, k_1: biodegradation rate

Design of experiments (DOE) or statistical factorial design is a systematic method to determine the relationship between factors that affect a process (i.e. response) and the output of that process i.e. a response variable (Annadurai *et al.*, 2000). In addition, this method also helps to rationalise and limit the number of experiments required to quantify the influence of particular factors on a response variable. This statistical method can evaluate the interactions among multiple factors that affect the biodegradation rate of chemicals and ultimately develop an empirical model based on the studied parameters to predict the biodegradation rate of

chemicals. Such experimental design techniques have been widely used to investigate the effect of different parameters on chemical removal rate and optimize different parameters to improve treatment efficiency, as mentioned in **Table 6-2** (Fannin *et al.*, 1981; Chen *et al.*, 2009; Rigas *et al.*, 2009; Agarry and Ogunleye, 2012). However, the use of factorial design to investigate the effect of initial chemical concentration and putative degrader numbers on the chemical removal rate in wastewater (i.e. activated sludge) has not been previously reported. The aim of this study was to investigate the influence of chemical concentration and starting putative degrader abundance on the biodegradation rate of aromatic chemicals in activated sludge using a factorial design approach. The ultimate overall goal was to develop an empirical model to predict biodegradation rate that could be interpreted in the context of established theories of biodegradation kinetics. Phenol and 2,4-dichlorophenol (2,4-DCP) were selected as model chemicals, as they possess different structural characteristics and are classified as readily and not readily biodegradable chemicals (EPA, 2011), respectively, and were also included in the previous chemical dataset for QSBR model development (**Chapter 3**). The qPCR assays [i.e. Phenol; PH & C23D and 2,4-DCP; *tfdB & tfdC* genes] developed previously (see **Chapter 5**) were used to quantify the key degraders for chemical biodegradation in an activated sludge inoculum containing a mixed microbial community.

Objectives

In order to accomplish the aforementioned aim, the following objectives were set:

a. To identify an appropriate approach to dilute the activated sludge grab samples in order to vary initial putative degrader numbers in an inoculum for batch biodegradation assays of phenol and 2,4-DCP
b. To use a Design of Experiment (DOE) approach to study the influence of initial chemical concentration and degrader numbers on the biodegradation rate of phenol and 2,4-DCP.
c. To develop and validate a statistical model to predict the biodegradation rate of these chemicals with chemical concentration and starting specific degrader abundance in inocula.

Table 6-2 Factorial design experiments in biodegradation studies

Factors	Response	Objective of Experiment	Reference
Phenol concentration, inocula concentration, mineral medium type, pH, flask closure, inoculum filtration	Phenol biodegradation rate	To study the main and interactive effects of six experimentally controlled environmental factors on phenol biodegradation	(Fannin *et al.*, 1981)
Combination of different bacterium concentration	TOC degradation rate	To establish a mixed culture inocula for enhanced wastewater treatment	(Chen *et al.*, 2009)
Straw content, Lindane content	Lindane degradation rate	To determine the optimum specific lindane degradation rate in soil by *Pleurotus ostreatus* with the aid of experimental design by optimizing the straw content and lindane content	(Rigas *et al.*, 2009)
NPK fertilizer amount, Tween 80 concentration, Mixed culture concentration	Weathered Bonny Light Crude Oil (WBLCO) removal	To evaluate the effect of biosimulation and bioaugmentation amendment agents (NPK fertilizer, Tween 80 and mixed culture) on the bioremediation of soil contaminated with crude oil using factorial design and optimized the best combination of these agents to achieve maximum crude oil removal.	(Agarry and Ogunleye, 2012)

6.3 Materials and Methods

6.3.1 Waste water treatment plant sampling

Activated sludge grab samples were collected from the biological aeration basin of four selected WWTPs operated by Northumbrian water Ltd in the North East of England as mentioned in Chapter 5, section 2.10.1.

6.3.2 Dilution of activated sludge

According to the results from a previous chapter (Chapter 5), the activated sludge grab sample from WWTP4 was only considered for the biodegradation assay of 2,4-DCP (since 2,4-DCP biodegradation, and detection of key catabolic genes involved its degradation, were only observed in this WWTP of those studied), whereas the phenol biodegradation assay was performed using activated sludge grab samples from all four WWTPs. Activated sludge grab samples were diluted to vary the initial competent degrader numbers in inocula. As a starting point, the activated sludge grab samples that were treated with ultraviolet (UV) light for 40 minutes were used to dilute the untreated activated sludge. The rationale behind the UV sterilization was primarily to inactivate the cells present in the activated sludge, and secondly to maintain identical matrices in both fresh sludge and the UV treated sludge used to dilute the fresh activated sludge. For UV treatment, a UV irradiation dose of 110 kJ/cm^2 was applied using an UV lamp (254 nm UV lamp T-514, Semtec Flow Water Sterilization, China) (Petropoulos, 2015). The activated sludge was passed through the chamber where the lamp

was situated using a peristaltic pump (530U/R, Watson-Marlow Limited, Cornwall, United Kingdom) at a flow rate of 2400 mL/min. The activated sludge was sampled at 0, 5, 10, 15, 20, 25, 35 and 40 mins, and total cell counts were quantified as mentioned elsewhere (Petropoulos, 2015) to assess the sterilization efficiency. Briefly, serial dilutions of samples (1 - 10^5) were prepared in sterile Phosphate Buffer Solution (PBS). 100 µL of diluted samples were inoculated in the R2A agar plates in and spread with sterile L-shaped plastic rod. Plates (dilution and controls [PBS was inoculated instead of sample]) were prepared in duplicate; additional un-inoculated R2A agar plates were also prepared as controls. All the plates were incubated at 22 ^oC in the dark for 3 days prior to enumeration of viable cells. The calculation of number of viable cells was based on APHA (American Public Health Association) (APHA, 2005). The appropriate dilution method was developed according to following steps:

1. The biodegradation assays for phenol (25mg/L) and 2,4-DCP (5 mg/L) were respectively performed with activated sludge grab samples from WWTP1 and WWTP4 using 40 minute UV treated and untreated sludge to evaluate the efficiency of UV inactivation of the microorganisms present in activated sludge. Each assay was performed in duplicate.
2. If significant removal of test chemicals was observed over the duration of experiments in the reactor with UV-treated activated sludge, sterile OECD mineral medium (OECD, 1992a) was used to dilute the activated sludge. Otherwise, the UV treated activated sludge was used to dilute the fresh activated sludge.

6.3.3 Dilution of activated sludge with UV treated activated sludge

The effect of UV sterilization on total viable cell counts of activated sludge (WWTP1) and phenol biodegradation is shown on **Figure 6-1** (A and B). UV sterilization resulted in a 2 log_{10} reduction in the total viable cell counts initially present in the activated sludge. Significantly a small amount of phenol degradation was observed in the UV-treated activated sludge over 5 hours, whereas in untreated sludge, 63% of phenol was removed. These results suggest that the majority of microorganisms in the UV-treated sludge were inactive over the duration of experiments. It is well accepted that UV disinfection works by transferring electromagnetic energy through the cell wall of a microorganism, which ultimately destroys genetic material (DNA and RNA) and retards their ability to reproduce (Matasci *et al.*, 1999; Hassen *et al.*, 2000).

2,4-DCP degradation was observed in the both UV-treated and untreated activated sludge from WWTP4 (**Figure 6-2**). The first order biodegradation rates for both the systems were

similar (**Table 6-3**). These results suggest that the putative 2,4-DCP degrading microorganisms were either resistant to, or initially deactivated by, UV irradiation treatment which might then have revived over time and resulted in efficient 2,4-DCP degradation. It must be noted that microorganisms can repair and reverse the destruction effects of UV through repair mechanisms (Matasci *et al.*, 1999). Therefore, activated sludge grab samples from WWTP4 for 2,4-DCP biodegradation studies were diluted using OECD recommended mineral medium (OECD, 1992a).

A) B)

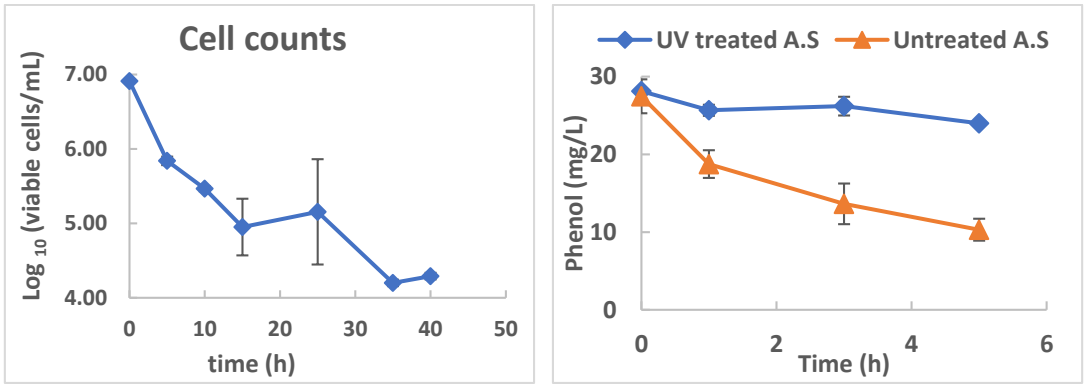


Figure 6-1Viable cell counts (on R2A agar plate) at different time intervals during UV sterilization of activated sludge grab sample (A), and phenol biodegradation using UV -treated and untreated activated sludge grab samples as inocula (B). Error bars in both figures are ± 1 s.d. of biological duplicate measurements.

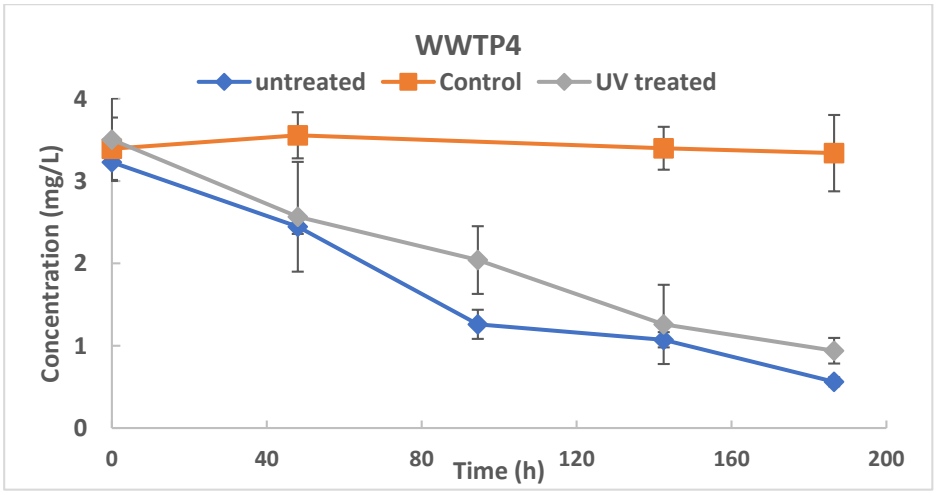


Figure 6-2. 2,4-DCP biodegradation in autoclaved (Control), UV-treated and untreated activated sludge. Experiments were conducted in duplicate with one control to check adsorption of 2,4-DCP on solids. For control experiment, a single reactor was run, while sampling was perfored in duplicate at each time point. Error bars are ± 1 s.d. of duplicate measurements.

Table 6-3 First order biodegradation rates of 2,4-DCP in UV-treated and untreated activated sludge.

A.S. inoculum source	UV Treatment	Average Rate (h^{-1})	s.d
WWTP4	yes	0.0072	0.000424
WWTP4	no	0.0092	0.000495

6.3.4 Experimental design for Biodegradation assay

A strategically planned and executed experiment will provide information about the effect on a response variable due to one or more factors. A factorial design 2^2 was used to study the influence of initial chemical concentration and competent degrader number on biodegradation of phenol and 2,4-DCP in activated sludge grab samples, as mentioned in the previous section. The initial concentrations of chemicals and specific degraders used in the batch experiments are reported in **Table 6-4**. The highest concentration used for both chemicals are not reported to show inhibitory effects on microbial growth (Tyler and Finn, 1974; Abuhamed *et al.*, 2004). Experiments were randomized and each performed in duplicate. A matrix was then established according to their high and low levels, represented by +1 and -1, respectively as shown in **Table 6-5**.

Table 6-4 Level of two variables in coded (+1/-1) and un-coded form for the degradation of phenol and 2,4-DCP using activated sludge (AS) grab samples as inocula. Two ten-fold different concentration ranges for the test chemicals are used in order to avoid possible inhibitory or toxic effects on microbial growth. The dilution of inocula used in phenol degradation assays were performed with UV-sterilized activated sludge, whereas inocula dilutions in 2,4-DCP assays were performed with OECD mineral medium.

Factors	High level (+1)	Low level (-1)
Initial chemical concentration (C_0)	25 mg/L (phenol) or 5 mg/L (2,4-DCP)	2.5 mg/L (phenol) or 0.5 mg/L (2,4-DCP)
Total catabolic gene copies/mL AS (X_0)	Undiluted AS grab sample	Diluted (1 in 10) AS grab sample

Table 6-5 Experimental conditions (for both 2,4-DCP and phenol assays) created with a factorial design (2^2) for two variables. Each experiment was performed in duplicate. The dilution of inocula used in phenol degradation assays were performed with UV-sterilized activated sludge (AS), whereas inocula dilutions in 2,4-DCP assays were performed with OECD mineral medium.

Run Order	Initial total catabolic genes/mL AS (X_0)	Initial spiked chemical concentration (C_0) [mg/L]
1	+1	-1
2	-1	+1
3	+1	+1
4	-1	+1
5	-1	-1
6	+1	+1
7	+1	-1
8	-1	-1

The statistical and graphical analysis was carried out using Minitab (Minitab17, 2010). In order to elucidate the relationship between the experimental variables (i.e. initial catabolic gene copies/mL of AS and initial spiked chemical concentration) and response (biodegradation rate), a multivariate regression model was developed and subsequently

validated with further degradation assays conducted with varied experimental variables as shown in **Table 6-6** and **Table 6-7**.

Table 6-6 Experimental conditions of reactors for phenol DOE validation experiments. The dilution of inocula were performed with UV-sterilized activated sludge (AS).

A.S source	ID	Initial catabolic gene copies/mL AS (X_0)	Initial spiked phenol concentration (C_0)
WWTP1	M	0.9 * X_0	5
	N	0.75 * X_0	15
WWTP3	X	0.6 * X_0	2.5
	Y	0.8 * X_0	10
	Z	0.9 * X_0	20
WWTP2	F	0.2 * X_0	7.5
	G	0.5 * X_0	10
	H	0.6 * X_0	25

Table 6-7 Experimental condition of reactors for 2,4-DCP DOE validation experiments. The dilutions of activated sludge (AS) inocula were performed with OECD mineral medium.

A.S source	ID	Initial catabolic gene copies/mL AS (X_0)	Initial spiked 2,4-DCP concentration (C_0)
WWTP4	X1	0.1 * X_0	2.5
	X2	0.1 * X_0	2.5
	Y1	X_0	2.5
	Y2	X_0	2.5
	Z1	0.5 * X_0	1
	Z2	0.5 * X_0	1

6.3.5 Biodegradation assays

Reactors for biodegradation assays were setup for the designed experimental conditions (training and validation experiments) as mentioned in **Table 6-5**, **Table 6-6** and **Table 6-7.** Activated sludge grab samples were used as inocula, and where required, dilution was performed either with UV-treated activated sludge (for 40 minutes) or OECD recommended mineral medium, according to the steps suggested in section 6.3.2. Control reactors for each test chemical were prepared with inactivated activated sludge [inactivation was performed either with UV treatment or by autoclaving activated sludge (autoclaved at 121 °C for 20 min and 103 kPa, two times, 24 h apart)(Helbling *et al.*, 2012), based on the method of dilution

used]. Biodegradation assays were conducted according to the method described in **Chapter 5, section 5.3.13**.

6.3.6 DNA extraction and gene quantification

Genomic DNA was extracted from the samples collected during the biodegradation assay using the FastDNA Spin Kit for Soil according to the manufacturer's instructions (MPBiomedicals, Santa Ana, CA, USA). The extracted DNA was stored at -20 ^{o}C until further use. The major catabolic genes associated with bacterial phenol and 2,4-DCP metabolism in the activated sludge were quantified with a real time quantitative PCR method. The PH and C23D genes, which respectively encode the PH and C23D enzymes, were used as biomarkers for putative phenol degraders. Whereas for 2,4-DCP, *tfdB* and *tfdC* genes (encoding 2,4-DCP hydroxylase and chlorocatechol-1,2-dioxygenase enzymes) were used as biomarkers for putative 2,4-DCP degraders. The active catabolic genes in reactors containing inoculum diluted with UV-treated sludge was calculated using following formula for each gene:

$A = (1 - p) * B$

Where,

A; the active catabolic gene concentration (genes copies /mL) in the AS inocula (untreated AS)

B; the total catabolic gene concentration (genes copies/mL) in the AS inocula (UV-treated AS + untreated AS)

p; fraction of UV-treated AS added for dilution as a proportion of the overall reactor contents

6.4 Results and Discussion

6.4.1 DOE results for phenol and 2,4-DCP biodegradation assays

For the majority of conditions, the biodegradation of phenol (**Figure 6-3** and **Figure 6-4**, training and validation experiments respectively) and 2,4-DCP (**Figure D 1 and Figure D 2, Appendix D**, training and validation experiments respectively) were better described by first-order kinetics than zero- or second-order kinetics **[Table D 1 and D 2 (Appendix D)].** The biodegradation of test chemicals followed a general trend that higher putative specific degrader concentrations exhibited greater rates of biodegradation than those with lower degrader concentrations at a given spiked chemical concentration. UV-treated activated sludge controls showed little or no biodegradation of phenol (**Figure 6-5)** thus verifying the efficacy of UV sterilization. Likewise, **Figure D 3 (Appendix D)** shows the biodegradation profile of 2,4-DCP in reactors with autoclaved activated sludge (different dilutions), where 2,4-DCP spiked at different concentrations remained constant over the duration of assays. Furthermore, the observed biodegradation profile in control assays

indicates that both phenol and 2,4-DCP degradation were brought about by biotic process (i.e. biological removal). However, approximately 75% of the spiked 2,4-DCP was measured at

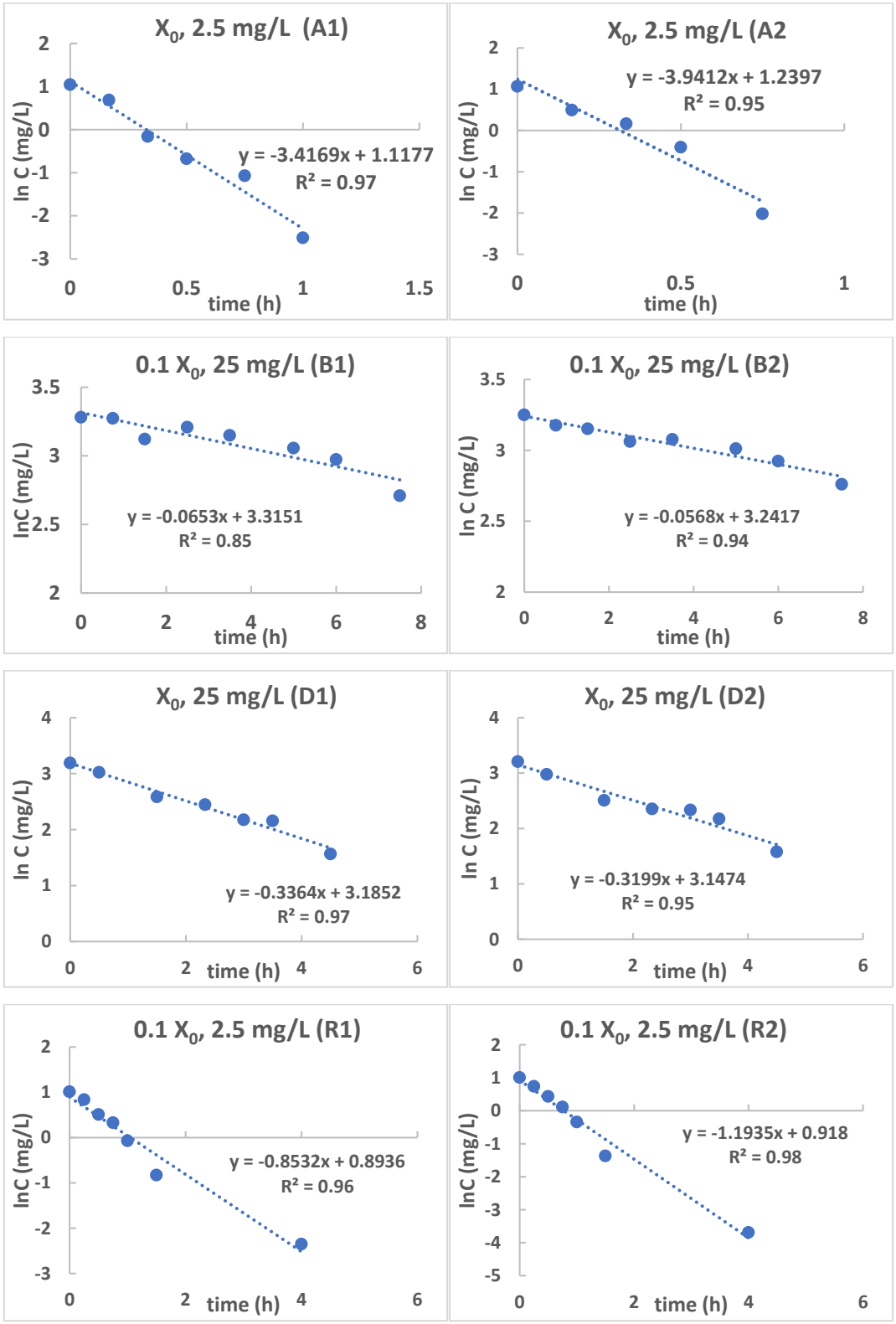


Figure 6-3 Phenol biodegradation profile for experimental conditions created with a factorial design (2^2) for the training experiments. The figure chart title shows the experimental condition of each reactor.

time zero in the liquid phase of the reactors containing activated sludge (both treatment and control), suggesting an instant adsorption of 2,4-DCP on the sludge surface. This was

confirmed by the initial measurements in the abiotic control reactors (i.e. in DI water), which showed 100% recovery at time zero (**Chapter 5, section 5.4.3.2**).

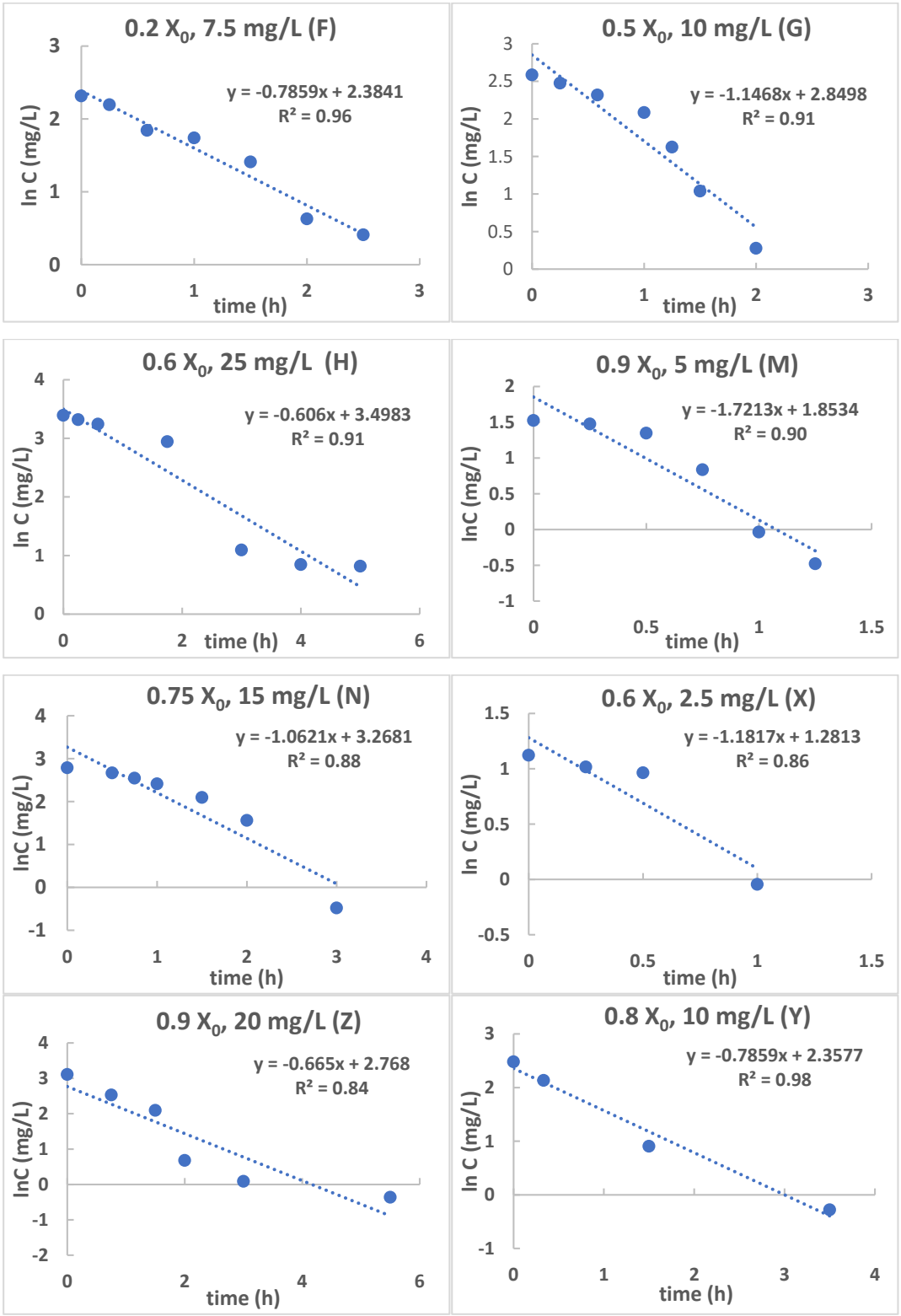


Figure 6-4 Phenol biodegradation profiles for DOE validation experiments. The figure chart title shows the experimental condition of each reactor.

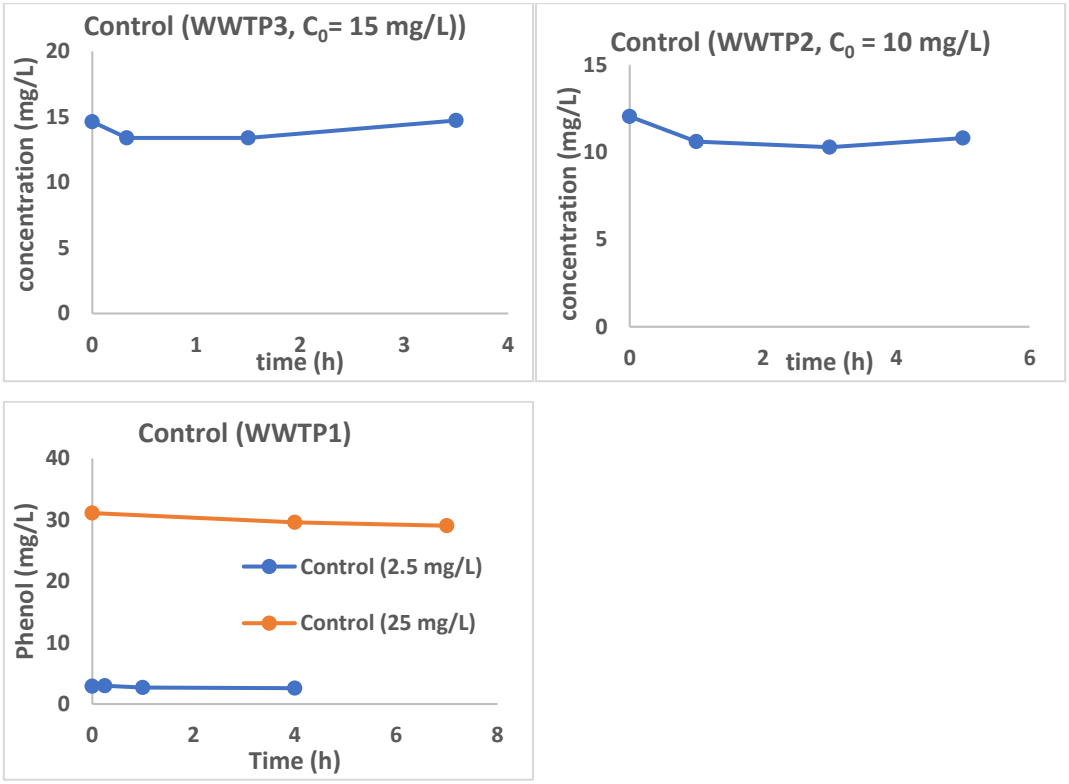


Figure 6-5 Phenol biodegradation profile in UV treated activated sludges from different WWTPs, which acted as a control reactor.

The initial total catabolic genes involved in phenol and 2,4-DCP biodegradation pathways were quantified and presented in **Table 6-8** and **Table 6-9**, respectively.

Table 6-8 Summary of the initial catabolic gene copy number per mL of activated sludge (AS) in different reactors used for phenol DOE and validation experiments, observed and model predicted phenol biodegradation rates for each reactor (Top; PH gene and Bottom; C23D gene). Gene copy numbers were quantified with qPCR using the method mentioned in Section 6.3.6. The number in parentheses represents the standard deviation for gene copy numbers.

Experiment Type	ID	X_0 (PH gene copies/mL AS)	Average X_0	C_0 (mg/L)	Experimental rate (h^{-1})	Predicted rate (h^{-1})
Training	A1	8.11×10^4	9.87×10^4 (1.74×10^4)	2.5	3.41	3.25
	A2	1.21×10^5		2.5	3.94	4.46
	D1	1.02×10^5		25	0.34	0.32
	D2	9.02×10^4		25	0.32	0.29
	B1	9.77×10^3	7.79×10^3 (1.5×10^3)	25	0.065	0.064
	B2	8.06×10^3		25	0.057	0.06
	R1	6.33×10^3		2.5	0.85	0.99
	R2	7.02×10^3		2.5	1.19	1.018
Validation	M	8.71×10^4		5	1.72	3.08
	N	7.57×10^4		15	1.062	1.51
	X	6.76×10^4		2.5	1.18	2.84
	Y	7.50×10^4		10	0.79	2.12
	Z	6.46×10^4		20	0.66	0.78
	F	1.67×10^5		7.5	0.78	1.57
	G	1.39×10^5		10	1.14	3.48
	H	1.57×10^5		25	0.60	0.48

Experiment Type	ID	X_0 (C23D gene copies/mL AS)	Average X_0	C_0	Experimental rate (h^{-1})	Predicted rate (h^{-1})
Training	A1	1.33×10^6	1.34×10^6 (1.06×10^5)	2.5	3.42	3.77
	A2	1.44×10^6		2.5	3.94	4.02
	D1	1.19×10^6		25	0.34	0.28
	D2	1.38×10^6		25	0.32	0.32
	B1	1.23×10^5	1.63×10^5 (2.94×10^4)	25	0.065	0.05
	B2	1.85×10^5		25	0.057	0.064
	R1	1.58×10^5		2.5	0.85	1.032
	R2	1.85×10^5		2.5	1.19	1.09
Validation	M	1.11×10^6		5	1.72	2.92
	N	9.88×10^5		15	1.06	1.45
	X	8.88×10^5		2.5	1.18	2.73
	Y	1.08×10^6		10	0.79	2.21
	Z	1.26×10^6		20	0.66	1.031
	F	4.91×10^5		7.5	0.78	1.43
	G	1.33×10^6		10	1.15	2.61
	H	1.37×10^6		25	0.60	0.32

X_0 : initial catabolic gene (PH or C23D) copies per mL of Activated sludge, C_0: spiked phenol concentration in reactors

Table 6-9 Summary of the initial catabolic gene copy numbers per mL of activated sludge (AS) in different reactors used for 2,4-DCP DOE experiments and validation experiments, observed and model predicted 2,4-DCP biodegradation rates for each reactor (Top; *tfdC* gene and bottom; *tfdB* gene). Gene copy numbers were quantified with qPCR using the method mentioned in Section 6.3.6. The number in parentheses represents the standard deviation for gene copy numbers.

Experiment Type	ID	X_0 (*tfdC* gene copies/mL A.S)	Average X_0	C_0	Experimental rate (h^{-1})	Predicted rate (h^{-1})
Training	M1	9.67E+05	8.54E+05 (7.52E+04)	0.5	0.11	0.143
	M2	8.23E+05		0.5	0.13	0.119
	N1	8.19E+05		5	0.0083	0.0081
	N2	8.09E+05		5	0.0085	0.008
	O1	1.33E+05	1.63E+05 (4.78E+04)	0.5	0.010	0.0156
	O2	2.04E+05		0.5	0.0075	0.0037
	P1	2.05E+05		5	0.0022	0.0029
	P2	1.12E+05		5	0.003	0.0021
Validation	X1	1.93E+05		2.5	0.0075	0.008
	X2	1.69E+05		2.5	0.0067	0.0066
	Y1	8.75E+05		2.5	0.034	0.075
	Y2	7.22E+05		2.5	0.031	0.06
	Z1	4.68E+05		1	0.030	0.054
	Z2	6.77E+05		1	0.025	0.085

Experiment Type	ID	X_0 (*tfdB* gene copies/mL A.S)	Average X_0	C_0	Experimental rate (h^{-1})	Predicted rate (h^{-1})
Training	M1	5.68E+05	4.56E+05 (1.30E+05)	0.5	0.113	0.156
	M2	5.07E+05		0.5	0.136	0.14
	N1	2.68E+05		5	0.0083	0.0057
	N2	4.82E+05		5	0.0085	0.0087
	O1	5.51E+04	5.19E+04 (1.99E+04)	0.5	0.0102	0.0097
	O2	2.31E+04		0.5	0.0075	0.0059
	P1	6.76E+04		5	0.0022	0.0028
	P2	6.18E+04		5	0.003	0.0027
Validation	X1	4.79E+04		2.5	0.0075	0.0054
	X2	6.08E+04		2.5	0.0067	0.0075
	Y1	2.85E+05		2.5	0.034	0.045
	Y2	2.71E+05		2.5	0.031	0.042
	Z1	2.87E+05		1	0.0301	0.069
	Z2	2.30E+05		1	0.025	0.054

X_0 : initial catabolic gene (*tfdC or tfdB*) copies per mL of Activated sludge, C_0: spiked 2,4-DCP concentration in the reactors

The putative chemical degrader's proportion to the overall biomass in the activated sludge inocula was estimated from the quantitative gene copy data obtained from the real-time PCR and the total bacterial cell count using flow cytometry. The percentage of putative phenol degraders in the inocula ranged from 0.0015 to 0.041% and 0.03 to 0.36% based on *PH* and *C23D* genes, respectively. Whereas, the percentage of putative 2,4-DCP degraders in the

activated sludge inocula ranged from 0.009 to 0.23% and 0.046 to 0.36% based on *tfdB* and *tfdC* genes, respectively. These results clearly indicate that bacteria having the metabolic capability to degrade specific organic pollutants (e.g., phenolic compounds) in the environment are present in low abundances. It should be noted that, multiples copies of a gene might be present in some bacterial species degrading xenobiotics (Pérez-Pantoja *et al.*, 2010). In order to evaluate this, *PH* gene copy numbers in *Pseudomonas sp.* CF600 was quantified with qPCR and compared with total cell counts enumerated with flow-cytometer. On average, each *Pseudomonas sp.* CF600 has 4.5 *PH* gene copies per cell (s.d +/- 1.18). Likewise, it has been verified elsewhere that metabolically versatile bacterial species can possess two or more copies of gene (Pérez-Pantoja *et al.*, 2010). Perez –Pantoja et al (2010) identified two copies of PH in *Cuprividus nectar JMP134*. These findings suggest that using gene copy number as an estimate of total putative degraders might potentially overestimates initial degrader's numbers in the inocula and ultimately have an influence in the model development and outcomes.

6.4.2 Multi-regression model

Based on the experimentally determined biodegradation rates, further analysis of the factorial design was performed using the Minitab (Minitab17, 2010). A multivariate regression model [see equations 1(A and B) and 2 (A and B) below] relating the response (i.e. first order biodegradation rate) and independent variables (catabolic gene copy numbers and initial chemical concentration) were estimated, the coefficient of each studied independent variables and their interaction term were determined [**Figure 7-2**, **Figure 7-3**, **Table D 3 (Appendix D) and Table D 4 (Appendix D)].**

The multivariate regression **Equations 1 (A and B) and 2 (A and B)** represent the mathematical models relating the first order biodegradation rate (*k*) of phenol and 2,4-DCP in activated sludge with independent process variables (PH, C23D, *tfdC, tfdB* abundances and spiked chemical concentration).

$$k_{phenol(PhH)} = 1.2996 + 0.749 * X_0 - 1.1119\, C_0 - 0.6204 * X_0 * C_0 \quad (1A)$$

$$k_{phenol(C23D)} = 1.2996 + 0.749 * X_0 - 1.1119\, C_0 - 0.6204 * X_0 * C_0 \quad (1B)$$

$$k_{2,4-DCP(tfdB)} = 0.03615 + 0.0304 * X_0 - 0.0306\, C_0 - 0.0275 * X_0 * C_0 \quad (2A)$$

$$k_{2,4-DCP(tfdC)} = 0.03615 + 0.0304 * X_0 - 0.0306\, C_0 - 0.0275 * X_0 * C_0 \quad (2B)$$

Where, X_0 and C_0 are the coded values of the test variables; initial catabolic gene [PH *(Eq 1A) /* C23D *(Eq 1B) / tfdB (Eq 2A) / tfdC (Eq 2B)*) copies/mL of AS and initial chemical concentration, respectively. Regardless of catabolic gene selected for each chemical (PH or

C23D in phenol and *tfdC or tfdC* in 2,4-DCP), the model equation to predict the first order rate of phenol and 2,4-DCP with the aforementioned genes are identical. However, it should be noted that, in the above equations, the coefficient of each test variable is a coded coefficient and will have different un-coded coefficients, which can be estimated according to Equation 3 (DOE, 2002).

$$X_{uncoded} = \frac{Hi+Lo}{2} + X_{coded}\ \frac{Hi-Lo}{2} \quad (3)$$

Where, $X_{uncoded}$ is the coded coefficient for test variable, *Hi* is the uncoded high level of the test variable, *Lo* is the uncoded low level of the test variable and X_{coded} is the uncoded coefficient for test variable

The above models can be used to predict the first order biodegradation rate of phenol and 2,4-DCP within the limits of the experimental factor values selected, which are given in **Table 6-8** and **Table 6-9**, respectively. **Figure 6-6 (A, B, C and D)** shows that the response values highly correlate with the model predicted first order rates for both chemicals ($p < 0.05$, R^2 = 0.70 [phenol with *PH* gene], 0.77 [phenol with C23D gene], 0.92 [2,4-DCP with *tfdB* gene] and 0.80 [2,4-DCP with *tfdC* gene]). The model predicted rates for the phenol and 2,4-DCP validation assays (with all four genes) were, in general, higher than the experimentally determined rates by an average factor of 1.89 (phenol with PH: 2.07, phenol with C23D: 1.98, 2,4-DCP with *tfdB*: 1.64 and *tfdC*: 1.89).

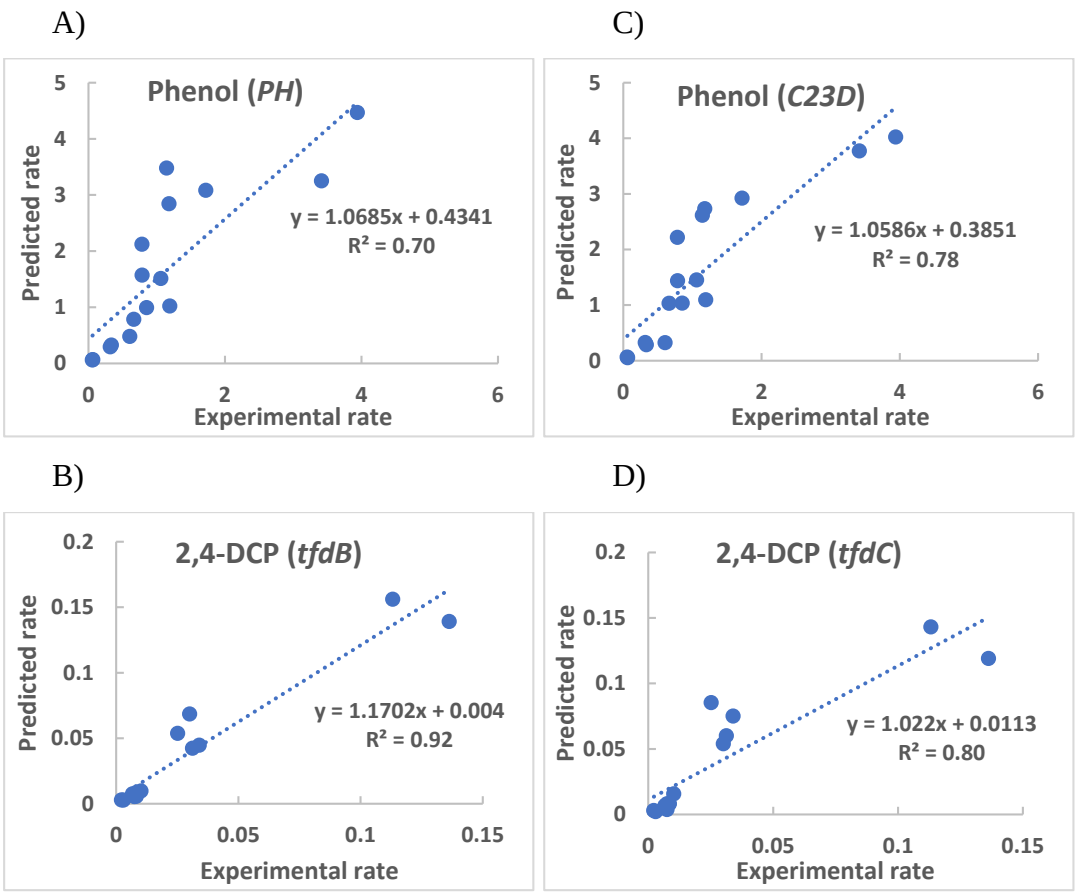


Figure 6-6 Relationship between the experimentally determined biodegradation rates and the model predicted biodegradation rates for (A) phenol with *PH* gene and (B) 2,4-DCP with *tfdB* gene, (C) phenol with *C23D* gene and (D) 2,4-DCP with *tfdC* gene.

Table 6-10 Analysis of variance for the multiple regression linear model (Equation 1A) for phenol from the factorial design experiments based of *PH* gene copy numbers.

A)

Source	**DF**	**Adjusted Sum of Squares**	**Adjusted Mean square**	**F-Value**	**p-Value**
Model	3	17.45	5.81	1783.88	0.000
Linear	2	14.37	7.18	2203.82	0.000
X_0	1	4.48	4.48	1375.73	0.000
C_0	1	9.88	9.88	3031.92	0.000
2-way interaction	1	3.08	3.08	943.99	0.000
X_0*C_0	1	3.08	3.08	943.99	0.000
Error	4	0.013	0.0033		
Total	7	17.47			

B)

Coded Coefficients					
Term	**Effect**	**Coefficient**	**SE Coefficient**	**T-Value**	**P-Value**
Constant		1.29	0.02	64.36	0.000
X_0	1.49	0.75	0.02	37.09	0.000
C_0	-2.22	-1.11	0.02	-55.06	0.000
X_0*C0	-1.24	-0.62	0.02	-30.72	0.000

The statistical significance of the four models were further tested through analysis of variance (ANOVA) presented in **Table 6-10 A (Phenol: PH gene), Table D3A [Appendix D] (Phenol:C23D gene), Table 6-11 A (2,4-DCP; *tfdB* gene) and Table D 4A [Appendix D] (2,4-DCP; *tfdC* gene)**. The ANOVA of the multivariate regression models indicate that the models were significant. In all models, it was found that the variables, X_0, C_0 and X_0*C_0, were significant model terms (p-value < 0.05). The predicted R^2 of 0.70 (Phenol; PH gene), 0.77 (Phenol; C23D gene), 0.92 (2,4-DCP; *tfdB* gene) and 0.80 (2,4-DCP; *tfdC* gene) were in reasonable agreement with the adjusted R^2 of 0.99 (phenol; both genes) and 0.98 (2,4-DCP; both genes), respectively. This implies that the prediction of experimental data is satisfactory. The fit of the model expressed by the coefficient of regression R^2 were found to be 0.99 and 0.99 respectively for phenol (both genes) and 2,4-DCP (both genes) models, which indicates that approximately 99 % of the variability in the rates could be explained by these models.

Table 6-11 Analysis of variance for multiple regression linear model (equation 2) for 2,4-DCP factorial design experiments based of *tfdB* gene copies number

A)

Source	**DF**	**Adjusted Sum of Squares**	**Adjusted Mean square**	**F-Value**	**p-Value**
Model	3	0.02	0.007	103.31	0
Linear	2	0.014	0.0075	110.2	0
X_0	1	0.0074	0.0074	109.39	0
C_0	1	0.0075	0.0075	111.01	0
2-way interaction	1	0.006	0.006	89.53	0.0001
X_0*C_0	1	0.006	0.006	89.53	0.0001
Error	4	0.0002	0.00006		
Total	7	0.002			

B)

Coded Coefficients					
Term	**Effect**	**Coefficient**	**SE Coefficient**	**T-Value**	**P-Value**
Constant		0.036	0.003	12.43	0
X_0	0.06	0.030	0.003	10.46	0
C_0	-0.061	-0.031	0.003	-10.54	0
X_0*C_0	-0.055	-0.027	0.003	-9.46	0.001

The coefficient of each terms included in models are shown in **Table 6-10 B (Phenol: PH gene), Table D 3B [Appendix D] (Phenol:C23D gene), Table 6-11 B (2,4-DCP; *tfdB* gene) and Table D 4B[Appendix D] (2,4-DCP; gene).** The initial catabolic gene copies/mL of AS (X_0) have a positive effect on chemical degradation rate whereas, initial chemical concentration (C_0) and the interaction between X_0 and C_0 have a negative effect on biodegradation rate. This implies that the biodegradation rates in reactors with higher starting specific degrader abundance for any given chemical concentration was higher than those with lower specific degrader abundance. In contrast, the degradation rate at a given starting specific degrader abundance in the reactor with lower initial chemical concentrations is higher than those with higher chemical concentrations. Similar observations have been made elsewhere (Thouand *et al.*, 1995; Berg and Nyholm, 1996). The probability of observing chemical biodegradation with environmental inocula increased with increasing specific degrader abundance in the inocula (Thouand *et al.*, 1995). In contrast, the extent of test chemical degradation by unadapted (i.e. no previous exposure to chemical) inocula tended to be higher with chemicals spiked at lower rather than higher concentrations (Berg and Nyholm, 1996), which might be due to an increased inhibitory effect on the microbial growth with

increasing chemical concentration. It has to be noted that the inocula used in the biodegradation assays in this study were not exposed to test chemicals prior to the experiment and are known to harbour putative chemical degraders. In the context of this study, the interaction term indicates the combined effect of chemical and putative degrader abundance on the extent of degradation.

6.4.3 Effect of initial chemical concentration and specific degrader numbers on rate

The factorial main effect plots (**Figure 6-7 A and C**) for both chemicals suggest that the biodegradation rate of chemicals increases with increasing initial degrader numbers, but decreases with increasing chemical concentration in a system. The effect of both these factors on biodegradation rate are significant. The interaction plots (**Figure 6-7 B and D**) further suggest that the initial degrader number has greater influence on biodegradation rate at lower chemical concentration than at higher concentration. These findings align with those parameters considered to be most important in, and captured by, theoretical models of biodegradation kinetics described in the literature (Simkins and Alexander, 1984; Simkins and Alexander, 1985; Okpokwasili and Nweke, 2006) (i.e. chemical degradation kinetic is associated with both initial substrate concentration and putative chemical degrader abundance). Concentration of test chemicals, the physico-chemical conditions and concentration of the bacterial inoculum are the three most important parameters that have significant influence on chemical biodegradability in regulatory biodegradation tests (Berg and Nyholm, 1996; Thouand *et al.*, 1996; Kowalczyk *et al.*, 2015b). Results from this study clearly support these observations. The observed differences in the rate of biodegradation across reactors appears to be mostly due to variation different parameters, among which the chemical concentration and initial putative degrader number among the reactors are the important **(Table 6-8 and Table 6-9).** Such variation might have an impact on the synthesis of catabolic enzymes in the bacterial cell which catalyse different reaction steps in the biodegradation pathway of chemicals (Grady, 1985; Kowalczyk *et al.*, 2015b). Alternatively, the time required for microbial adaptation to the chemicals may also vary with applied chemical concentration (Berg and Nyholm, 1996).

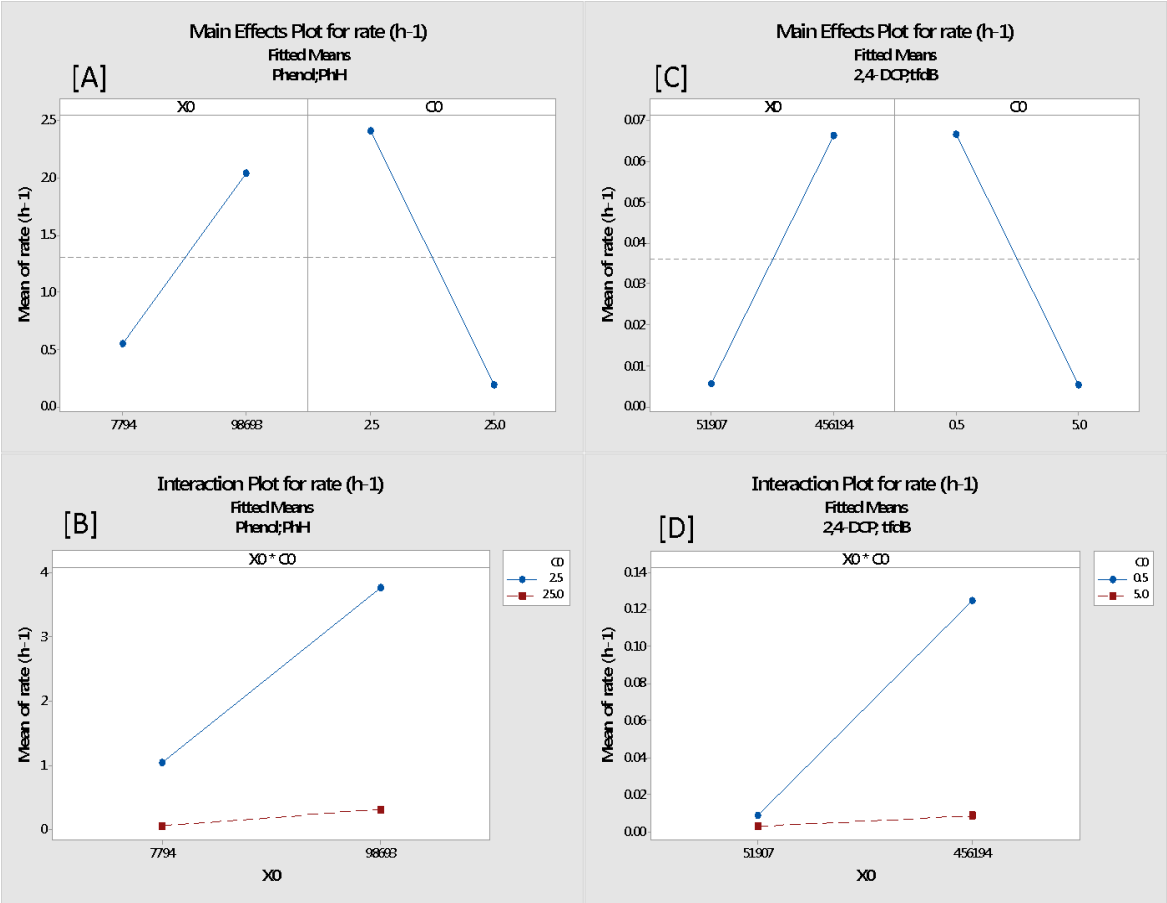


Figure 6-7 Main effect (A and C) and interaction plots (B and D) for first order biodegradation rates in relation to intial phenol (with *PH* gene) and 2,4-DCP (*tfdB* gene) degrader concentrations.

Table 6-12 Correlation analysis between the first order biodegradation rate of test chemicals with different parameters (X_0, C_0 and X_0/C_0)

Chemicals	Parameters	Pearson Correlation Coefficient (r)	P-value
2,4-DCP	C_0	-0.525	0.054
	X_0 (*tfdB*)	0.758	0.002
	X_0 (*tfdC*)	0.604	0.022
	X_0(*tfdC*)/C_0	0.953	0
	X_0(*tfdB*)/C_0	0.97	0
Phenol	C_0	-0.685	0.003
	X_0 (*C23D*)	0.418	0.107
	X_0 (*PH*)	0.241	0.37
	X_0(*C23D*)/C_0	0.937	0
	X_0(*PH*)/C_0	0.879	0

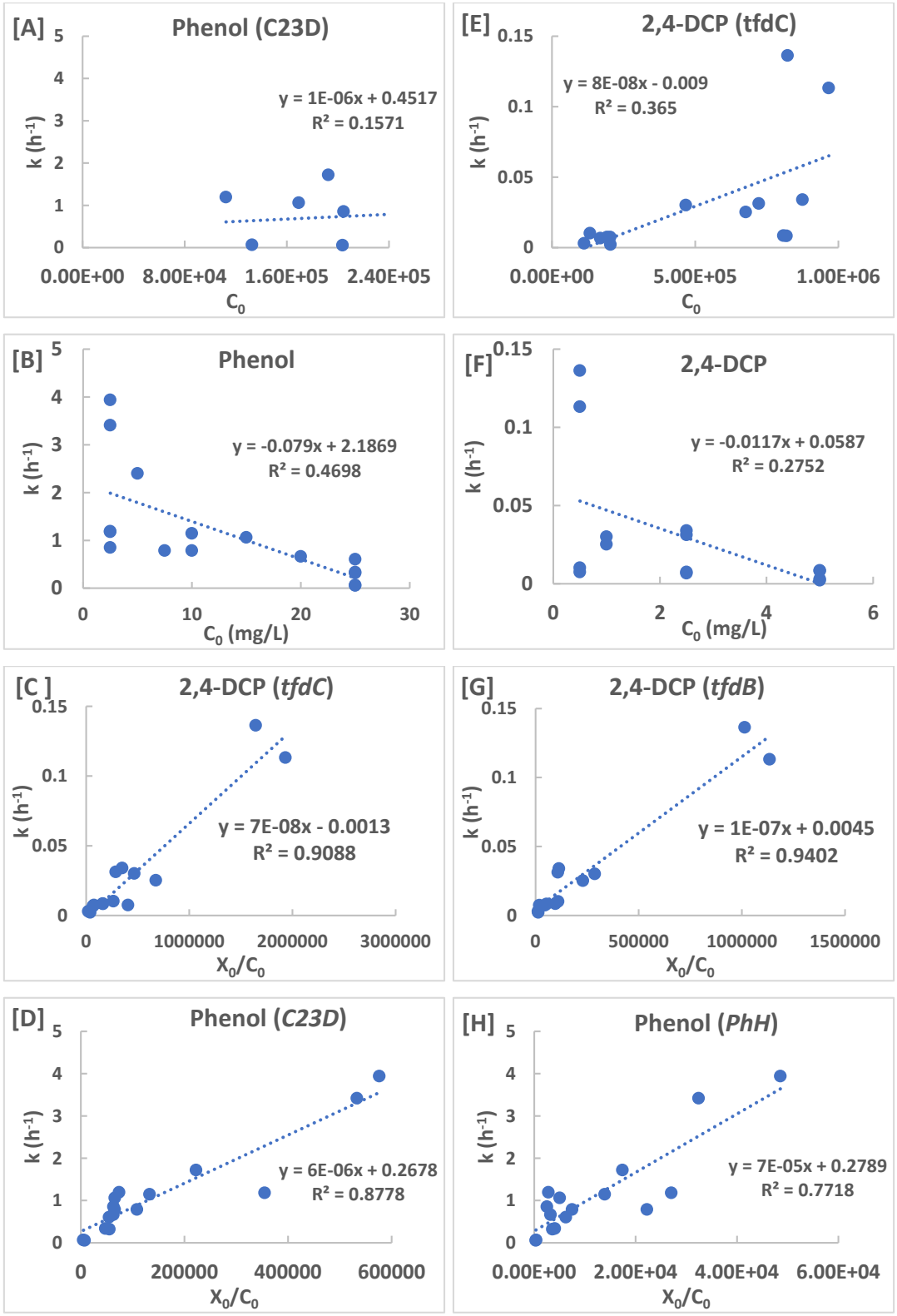


Figure 6-8 Relationship between first order biodegradation rates (*k*) of test chemicals (phenol and 2,4-DCP) with different parameters (X_0, C_0 and X_0/C_0). A and E: first order biodegradation rate and initial catabolic gene copies per mL of activated sludge (X_0); B and F: first order biodegradation rate and initial spiked test chemical concentration (C_0); and C, D, G and H; first order biodegradation rate and ratio of X_0 and C_0. The figure chart title shows the details (i.e. spiked chemical and gene involved) of each analysis.

The interaction term (i.e. $C_0 * X_0$) in the model has significant influence on biodegradation rate, indicating that the experimentally estimated biodegradation rates of the chemicals are a function of these two parameters. To further verify this, individual correlation analysis was

performed between experimentally determined first order rates, initial chemical concentration, and initial putative chemical degrader numbers (i.e. catabolic gene copy number) in an inoculum. Strong and significant correlations [Pearson correlation coefficient (r) > 0.879 and p-value < 0.05] were observed between biodegradation rates and the ratio of X_0 and C_0 (i.e. X_0/C_0) for both chemicals studied (**Table 6-12, and Figure 6-8 C, D, G and H**). In contrast, no strong correlation was observed between biodegradation rates and initial chemical concentration and putative degrader numbers for both chemicals. These results clearly suggest that biodegradation of chemical in an environment should be viewed with respect to both intrinsic properties of chemicals (e.g., chemical concentration) and inocula characteristics (i.e. probability of encountering putative degraders of the test chemical within the test inocula). Therefore, understanding the occurrence and abundance of putative chemical degraders and chemical concentrations will ultimately enable in better predictions of biodegradation. The modelling approach used in this study further demonstrated the feasibility of applying key catabolic genes involved in chemical biodegradation in the prediction of chemical biodegradation rates in environmental systems.

6.5 **Conclusion**

Two multiple-regression models, which were based on individual catabolic genes (phenol: PH and C23D & 2,4-DCP: *tfdB* and *tfdC*) and spiked chemical concentration, were developed for phenol and 2,4-DCP to predict the biodegradation. The analysis of all the developed models highlighted that the starting abundance of putative degraders and the chemical concentration collectively determines the rate of chemical removal in the environment. Analysis of the biodegradation data further confirmed that biodegradation is a function of an interaction of both initial chemical concentration and putative degrader number in an environment due to the strong and significant linear correlation between rate and ratio of X_0 and C_0 (i.e. X_0/C_0). Therefore, understanding the occurrence and abundance of putative degraders and chemical concentrations in an environmental system will ultimately enable better prediction of biodegradation rate.

Chapter 7
Use of 16S rRNA amplicon sequencing and functional gene analyses to reveal the associations between putative degraders and chemical degradation in activated sludge.

Chapter 7. Use of 16S rRNA amplicon sequencing and functional gene analyses to reveal the associations between putative degraders and chemical degradation in activated sludge.

7.1 Abstract

Putative chemical degraders are considered as the key taxa in the microbial communities performing the biotransformation of chemicals in natural or engineered ecosystems. This work focuses on the use of 16S rRNA amplicon sequencing and functional gene analyses to reveal the association between putative competent degraders and chemical degradation in activated sludge. Biodegradation assays for test chemicals (phenol, m-cresol, 4-chlorophenol and 2,4-dichlorophenol) were performed individually at two different test chemical (5 mg/L and 2.5 mg/L) and inocula concentrations (original and one-tenth dilution). Degenerate primer-sets targeting the major catabolic genes involved in the biodegradation of test chemicals were designed or adopted from the literature, and were used to quantify these genes in the inoculum at the beginning and end of biodegradation assays using quantitative polymerase chain reaction (qPCR). Characterisation of microbial taxa in the inocula (only spiked with 5 mg/L of test chemicals) at the beginning and end of biodegradation assays were determined using 16S rRNA amplicon sequencing, while total biomass in the inocula was enumerated with flow cytometry. A few of the identified taxa that were enriched in the microbial communities during biodegradation were previously associated with the catabolic genes selected for functional gene analyses, indicating the presence of probable putative chemical degraders in the inocula under study. Results further show that both major catabolic genes, and by association putative degraders, significantly increased over the duration of assays, indicating the involvement of such degraders in the biodegradation of the test chemicals under study. Those enriched taxa putatively identified as potential degraders, in both diluted and undiluted inocula, were rare within the microbial community, and accounted for only 7.69 (s.d ± 5.5) and 11 (s.d ± 8) percentage of the total cell, respectively. In addition their enrichment did not necessarily influence the overall biomass count of the system. Application of 16s RNA amplicon sequencing and functional gene analyses techniques allowed the investigation of putative catabolic genes and taxa and their relationship with chemical biodegradation. Further such studies will enable better understanding of biodegradation outcomes.

7.2 Introduction

Previous biodegradation assays with pure cultures of specific degraders in minimal media clarified that observed biodegradation rates of test chemicals could be attributed to differences in the chemical structure, which resulted in varying levels of susceptibility to microbial attack

(Chapter 4). However, in natural and engineered systems, which are more complex matrices comprised of a diverse mixed microbial population, biodegradation of a chemical is a function of multiple factors, which includes both inherent (e.g., chemical structure and concentration) and external factors (e.g., inocula characteristics and environmental conditions) (Pavan and Worth, 2008). The reproducibility of regulatory biodegradation tests is a significant challenge as microbial inocula used in such tests are poorly prescribed, poorly controlled (Kowalczyk *et al.*, 2015b), and are uncharacterised. The observed high variability in regulatory chemical biodegradation test results is potentially due to differences in inocula characteristics, which determines the presence or absence of putative chemical degraders that may influence the biodegradation outcome and kinetics. The use of inocula at concentrations that are orders of magnitude lower than those in the environment, results in a low probability of their inclusion and hence variable biodegradation (Thouand *et al.*, 1995; Goodhead *et al.*, 2014; Martin *et al.*, 2017b).

Biodegradation is often viewed an intrinsic property of chemicals, while measurement of the presence and activity of potential degraders in microbial communities receives little practical attention when investigating the degradation of specific chemicals by a mixed microbial community, such as those in regulatory biodegradation tests. Therefore, there is a need to further understand the role that putative chemical degraders play in the biodegradation of chemicals in such tests. The challenge is that putative degraders are part of a diverse mixed microbial community.

Until recently, efforts to address this challenging task have been hampered by the lack of availability of suitable methods and theories to characterize the microbial diversity of the inocula in relation to biodegradation performance. In the last two decades, there has been a revolution in molecular based techniques (DNA and RNA) (Widada *et al.*, 2002; Cébron *et al.*, 2008; Helbling *et al.*, 2012; Kowalczyk *et al.*, 2015a; Meynet *et al.*, 2015) that has led to strategies for monitoring, discovering and identifying the abundance and diversity of both specific degraders and total microbial communities, and their activities during the degradation of xenobiotics. Yet, there have been few previous reports of their use in assessing chemical biodegradation tests (Head *et al.*, 1998; Helbling *et al.*, 2012; Goodhead *et al.*, 2014; Kowalczyk *et al.*, 2015a). Molecular microbial ecology techniques could therefore be used to gain such an understanding; to determine not only the overall size of the inoculum, but also its quality (number of specific degraders), and the diversity of functional communities involved in the biodegradation process.

Results from Chapter 3 and 4 highlighted the existence of relationships between chemical structure and biodegradation rate, whereas results of Chapter 5 revealed that chemical

degradation might be associated with the proliferation of key putative specific degraders, a phenomenon that might bring about their enrichment and subsequent shifts in the bacterial community composition within inocula containing a mixed bacterial community. Finally, Chapter 6 demonstrated that the biodegradation rate of chemical in an environment is a function of both initial chemical concentration and the abundance of chemical degraders in an inoculum. Microbial degradation kinetics have mostly been validated using chemical degradation information and rarely give insights on how chemical degraders may change during chemical degradation and their ultimate influence on degradation kinetics. Characterization of the total microbial community and quantification of key catabolic genes in the inocula will help to evaluate the relationship between microbial diversity and biodegradation outcomes, by considering the occurrence and abundance of putative degraders and, the shift and enrichment of taxa involved in the degradation of chemicals. The aim of this study was to use 16S RNA amplicon sequencing and quantification of functional gene abundances to understand the effect of chemical concentration, structure and inocula characteristics on the biodegradation rate of chemicals and the subsequent bacterial community response of enriched taxa. Phenol, m-cresol, 4-chlorophenol and 2,4-dichlorophenol (2,4-DCP) were selected as model chemicals, as their biodegradability in natural and engineered ecosystem are different (EPA, 2011), and also included in our previous chemical dataset for QSBR model development (Chapter 3). In addition, these chemicals are known to be degraded by diverse types of bacteria, such as *Pseudomonas, Ralstonia, Pseudoxanthomonas*, *Acinetobacter*, and *Burkholderia* (Assinder and Williams, 1990; van Beilen *et al.*, 1994; Steinle *et al.*, 1998; Shaw *et al.*, 1999; Kanaly *et al.*, 2002; Louie *et al.*, 2002; Palleroni *et al.*, 2010; Mukherjee and Bordoloi, 2012; Choi *et al.*, 2013; Paisio *et al.*, 2014; Liu *et al.*, 2016; Chen *et al.*, 2017).

Objectives

In order to accomplish the aforementioned aim, the following objectives were set:

a. To conduct the biodegradation experiments for four chemicals (phenol, m-cresol, 4-chlorophenol and 2,4-DCP) with varying initial concentration and degrader numbers in grab activated sludge inocula capable of efficiently degrading test chemicals and determine the respective biodegradation rates. The biodegradation assay results for 2,4-DCP from the previous chapter (Chapter 6.1) were used.
b. To conduct 16S amplicon sequencing and quantification of key functional genes to determine any variation in the activated sludge microbial community structure over the duration of chemical biodegradation experiments that may relate to the rate and enrichment of specific taxa putatively involved in chemical degradation.

7.3 Material and Methods

7.3.1 Chemicals, activated sludge sampling and dilution

Four chemicals were selected (phenol, m-cresol, 4-chlorophenol and 2,4-dichlorophenol) for biodegradation assays in this study. These chemicals were the same as those used is Chapter 4 and all were purchased from Sigma-Aldrich at > 99% purity (Dorset, United Kingdom). Based on the previous results (from **chapter 5 [sections 5.4.3**], i.e. activated sludge grab sample from WWTP4 degraded both phenol and 2,4-DCP when used as an inoculum), activated sludge grab sample from WWTP4 was considered as an inoculum for the biodegradation assays. Activated sludge grab samples were diluted with sterile OECD-recommended mineral medium (OECD, 1992a) to vary the initial number of competent degraders in inocula.

7.3.2 Batch biodegradation experiments

Biodegradation assays for each test chemical were performed according the experimental conditions shown in **Table 7-1**. Biodegradation assays for m-cresol, 4-chlorophenol and phenol were conducted together, while assays for 2,4-DCP was performed a week earlier. Fresh activated sludge grab samples from WWTP4 were used as inocula for both sets of experiments. Where required, dilution of inocula was performed in OECD recommended mineral medium. Duplicate reactors were setup for each condition. Control reactors for each test chemical were prepared with activated sludge that was autoclaved (121 °C for 20 min and 103 kPa) two times, 24 h apart (Helbling *et al.*, 2012). Biodegradation assays were conducted according to the method described in **Chapter 5, section 5.3.13**. Each reactor were inoculated with 200 mL activated sludge (diluted or undiluted) and spiked with a concentration of test chemical as mentioned in **Table 7-1**. The time zero sample for DNA extraction was collected prior to the spiking of the chemicals into the reactors. 0.75 mL of sample at the beginning and end of the biodegradation assays were fixed in an equivalent volume of 100 % ethanol, stored at -20 ^{o}C and later analysed by flow-cytometry (Becton Dickinson, California) to quantify the total cell count according to the method mentioned in section 5.3.14, Chapter 5.

Table 7-1 Experimental condition for the biodegradation assays of four test chemicals (phenol, m-cresol, 4-chlorophenol and 2,4-DCP)

Activated sludge source	Initial catabolic gene copies/mL activated sludge (X_0)	Initial spiked test chemicals concentration (C_0) [mg/L]
WWTP4	0.1 * X_0	2.5
	0.1 * X_0	5
	X_0	2.5
	X_0	5

7.3.3 DNA extraction, total catabolic gene quantification and PCR amplification (library preparation)

Total DNA from the samples (0.5 mL activated sludge inoculum) during the biodegradation assay was extracted using the FastDNA Spin Kit for Soil according to the manufacturer's instructions (MPBiomedicals, Santa Ana, CA, USA). The extracted DNA was stored at -20 ^{o}C until further use. The major catabolic genes associated with bacterial phenol, m-cresol, 4-chlorophenol and 2,4-DCP metabolism (Gao *et al.*, 2010) were quantified with a real time quantitative PCR method (**see section 5.3.8 and 5.3.9** for details, **Chapter 5**). The PH and C23D genes, which respectively encode the PH and C23D enzymes, were used as biomarkers for the phenol biodegradation assays. The C23D gene was used as a biomarker for m-cresol assays. Likewise, *tfdB* and *tfdC* genes (encode 2,4-DCP hydroxylase and chlorocatechol-1,2-dioxygenase enzymes) were used as a biomarker genes for 2,4-DCP assays, and *tfdB* gene for 4-chlorophenol biodegradation assays.

7.3.4 PCR amplicon library preparation

The DNA extracted from the reactor samples were used for amplicon library preparation. The amplicon library for next generation sequencing (Ion Torrent) was generated using a fusion-primer PCR method. The fusion primers consist of the Ion Torrent adaptor sequence and the primer sequence. In addition, the forward primers also contain a unique barcode sequence that allows the differentiation of the sequence corresponding to each sample during data analysis. Briefly, the samples were labelled with a unique 12 bp Golay barcode, added to the 5' end of the forward primer through a GAT spacer, and attached to the Ion Torrent adapter A (5'-CCATCTCATCCCTGCGTGTCTCCGACTCAG-3'), while the reverse primers were attached to the Ion Torrent adapter trP1 (5'- CCTCTCTATGGGCAGTCGGTGAT-3'). The primer-set [515f (5'-GTGNCAGCMGCCGCGGTAA-3') and 926r (5'-CCGYCAATTYMTTTRAGTTT-3')] were used to amplify the V4 and most of the V5 regions of the 16S rRNA gene (Vignola *et al.*, 2017). The PCR reactions were performed in a total volume of 25 µL, containing 0.5 µM of each universal primer, 1 µl DNA template (or sterile molecular grade water, for the negative control), 0.2 mM of dNTPs, 1 U Taq DNA polymerase, (Invitrogen, Life Technologies, Paisley, UK), 1.5 mM $MgCl_2$ and 2.5 ul of 10 x PCR buffer (minus Mg). The following thermocycler program was used for the amplifications: a denaturation cycle of 94^{o}C for 3 minutes followed by 30 cycles of denaturation at 94^{o}C for 45 seconds, 30 seconds at annealing temperature (T_a =55 ^{o}C) and elongation at 72 ^{o}C for 1 minute, with a final elongation step at 72 ^{o}C for 10 minutes. The PCR products were stored at -20^{o}C until further use.

PCR amplicons were cleaned and size selected using double-sided solid-phase reversible immobilisation (SPRI) beads (Agencourt AMPure XP system, Beckman Coulter) following the manufacturer's instructions and quantified using a Qubit® 2.0 Fluorometer following the manufacturer's protocol. Clean PCR amplicons from samples were also pooled in equimolar concentrations and sequenced on an Ion Torrent Personal Genome Machine (PGM) using a 316 ion chip. Amplicons were pooled and sequenced by Dr Amy Bell in the School of Engineering, Environmental Engineering research group, Newcastle University.

7.3.5 Microbial community structure analysis

The reads were filtered for quality (filtering criteria: perfect match to sequence barcode/primer, 200bp minimum sequencing length), and sequencing errors were subsequently detected and corrected using Quantitative Insight Into Microbial Ecology (QIIME) Denoiser (Reeder and Knight, 2010). The obtained 16S rRNA gene data were finally processed using the QIIME 1.9.1 pipeline (Caporaso *et al.*, 2010) (http://www.qiime.org) using default parameters unless otherwise noted. Briefly, the sequences were binned into OTUs using 97% identity threshold, and the most abundant sequence from each OTU was selected as a representative sequence for that OTU. Taxonomy was assigned to bacterial OTUs against a subset of Greengene database (http://greengenes.lbl.gov/) (DeSantis *et al.*, 2006). OTU representative sequences were aligned using UCLUST (Edgar, 2010). In total 398,533 bacterial 16S rRNA gene sequences were obtained, with between 16,218 and 51,428 sequence reads per sample. Because an even depth of sampling is required for alpha and beta diversity comparison (Shaw *et al.*, 2008), sequences were randomly rarified to 16,000 per sample for downstream analysis. For beta-diversity analysis, the Bray Curtis dissimilarity metric was calculated for the genus level OTU table and visualized using 2-dimensional non-metric multidimensional scaling plot (NMDS) (Clarke and Warwick, 2001) using Primer6.

7.3.6 Statistical Analysis

All statistical analyses were performed in Minitab 17 Statistical software (Minitab17, 2010). A two-sample t-test was performed to compare the initial and final catabolic gene copy concentrations and total cell counts in the activated sludge grab sample inocula at the beginning and end of the biodegradation assay for the individual chemicals. A one-way Analysis of Variance (ANOVA) was performed to compare the biodegradation rates of each chemical for the different studied conditions. A comparison of the relative abundances of identified bacterial groups at the beginning and end of the experiments for individual chemicals was conducted using STAMP (Parks *et al.*, 2014) for significant effect (G-test, $p<0.05$). Absolute abundance of putative chemical degraders (or genera), whose relative

abundance increased significantly over the duration of experiment, was estimated. The relative abundance of those genera was multiplied by the total bacterial cell count (express in cells/mL) measured with flow cytometer.

7.4 Result and Discussion

7.4.1 Summary of biodegradation assay results for test chemicals

All the test chemicals were effectively removed by both undiluted and diluted activated sludge at the spiked chemical concentrations (i.e. 2.5 mg/L and 5 mg/L). The degradation of these chemicals was best explained by first order kinetics. The biodegradation rates of each chemical for the different studied conditions were significantly different (one-way ANOVA, $p < 0.05$; **Table 7-2**). Likewise, for all the studied conditions, the estimated biodegradation rates were significantly different between the different test chemicals (one-way ANOVA, $p < 0.05$). Strong and significant correlations [Pearson correlation coefficient (r) ≥ 0.876 and p-value < 0.05] were observed between biodegradation rates and the ratio of X_0 and C_0 (i.e. X_0/C_0) for all test chemicals (**Figure 7-1**). The summary statistics of the regression analysis are presented in **Table 7-3**. The pattern of first order chemical removal rates for the test chemicals spiked at 5 mg/L in both diluted and undiluted inocula was phenol > m-cresol > 4-chlorophenol > 2,4-dichlorophenol. In reactors spiked with 2.5 mg/L of each of the test chemicals, the pattern was m-cresol > phenol > 4-chlorophenol > 2,4-dichlorophenol. The possible explanations for the observed differences in the chemical removal rate might be due to the difference in type of substituent groups attached in the aromatic ring, as substituent groups are associated with electronic properties of chemicals and influence their biodegradability (Pitter and Chudoba, 1990; Martin *et al.*, 2017c).

Table 7-2. Experimentally estimated biodegradation rates of four test chemicals for the selected experimental conditions. The numbere in parentheses represent the standard error of the slope (i.e. first order rate).

	5 mg/L, Diluted Inocula		**5 mg/L, Undiluted Inocula**		**2.5 mg/L, Diluted Inocula**		**2.5 mg/L, Undiluted Inocula**	
Chemicals	**1st order rate**	**s.d**	**1st order rate**	**s.d**	**1st order rate**	**s.d**	**1st order rate**	**s.d**
Phenol	0.095 (0.013)	0.024	0.75 (0.065)	0.0309	0.149 (0.005)	0.041	1.006 (0.119)	0.115
m-cresol	0.094 (0.005)	0.037	0.65 (0.067)	0.1241	0.168 (0.013)	0.089	1.13 (0.114)	0.135
4-CP	0.023 (0.002)	0.0028	0.039 (0.004)	0.0062	0.034 (0.001)	0.009	0.071 (0.005)	0.014
2,4-DCP	0.0026 (0.00001)	0.0005	0.008 (0.001)	0.0002	0.007 (0.0003)	0.0005	0.033 (0.004)	0.002

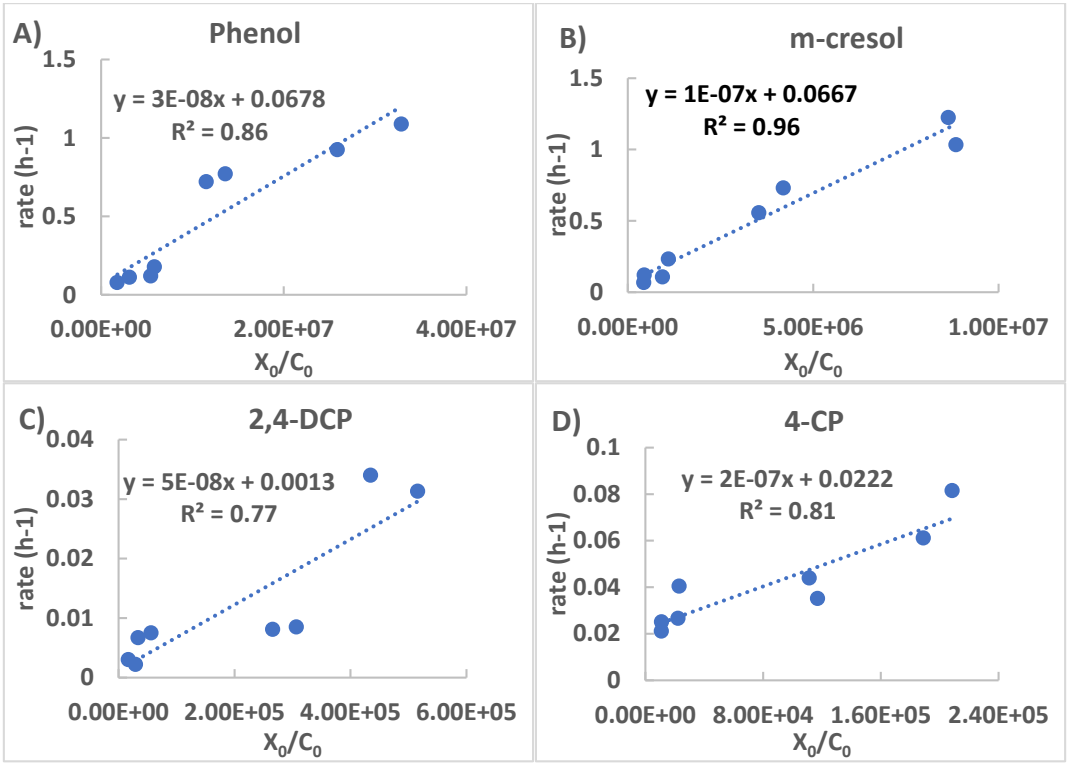


Figure 7-1 Relationship between first order biodegradation rates of test chemicals and the ratio of X_0 (initial putative competent degrader numbers) and C_0 (initial spiked chemical concentration). 2,4-DCP and 4-CP represent 2,4-dichlorophenol and 4-chlorophenol respectively.

Table 7-3 Summary statistics of univariate regression analysis between first order biodegradation rates of test chemicals and the ratio of X_0 (initial putative competent degrader numbers) and C_0 (initial spiked chemical concentration). 2,4-DCP and 4-CP represent 2,4-dichlorophenol and 4-chlorophenol respectively.

Chemical	Phenol	m-cresol	4-CP	2,4-DCP
R^2	0.86	0.96	0.816	0.766
Residual Std. Error	0.167	0.097	0.009	0.0065
Degrees of Freedom	7	7	7	7
P-value	8.36E-04	2.01E-05	2.00E-03	4.30E-03

7.4.2 Variation in functional gene and total bacterial abundances during test chemical biodegradation

Quantification of functional gene copy numbers

The results from previous **chapter 5 and chapter 6.1** confirmed the degradation capability of the inocula against the test chemicals used in this study. Therefore, the genes related to test chemical biodegradation, such as PH (phenol), C23D (phenol and m-cresol), *tfdB* (4-chlorophenol and 2,4-DCP) and *tfdC* (2,4-DCP) were quantified to estimate the putative chemical degraders in the inocula.

Real time quantitative PCR revealed the presence of major catabolic genes involved in the degradation of phenol, m-cresol, 4-chlorophenol and 2,4-DPC at all the time points (**Figure 7-2**), but the final catabolic gene copy numbers appeared to be significantly (p-value < 0.05,

2-sample t-test) greater than the initial ones (**Figure 7-2**) in all the test chemical degradation assays. The final abundance of studied catabolic genes in inocula spiked with 5 mg/ L of test chemicals were significantly higher than in inocula spiked with lower concentrations (i.e. 2.5 mg/L) (p-value < 0.05, 2-sample t-test), regardless of the dilution factor of the initial inoculum. This could be attributed to higher growth rate of specific degraders at the higher spiked concentration (i.e. 5 mg/L). The specific growth rate of bacteria at low concentration is a strong function of the substrate concentration, a result previously verified elsewhere (Shehata and Marr, 1971).

It should be noted that in biodegradation assays performed with diluted inocula (10 fold dilution), spiked with 5 mg/L and 2.5 mg/L of 2,4-DCP, the final relative abundance of catabolic genes were approximately 21- and 18-fold higher than the initial abundance respectively, while in the case of other chemicals were in the range of 1.7 to 5.3-fold higher (**Figure 7-2**). There might be multiple explanations for this difference. First, this might be mostly attributed to the reduction in the microbial diversity caused by dilution and addition of 2,4-DCP, which is toxic to many microbes not able to degrade 2,4-DCP in inocula, and might have provided suitable conditions for the growth of bacteria capable of degrading 2,4-DCP. Reduction in microbial diversity by dilution of the initial inoculum, or with addition of xenobiotics, or both, have previously been shown (Sutton *et al.*, 2013; Jung *et al.*, 2016). Addition of xenobiotics creates toxic effects on many microbes not able to metabolise 2,4-DCP within the communities, and the level of toxicity varies with biodegradability of added chemicals (Rücker and Kümmerer, 2012). This suggests that persistent chemicals might be toxic to a broad range of microbes as compared to readily biodegradable chemicals. It is also widely accepted that only a narrow group of specialized species are present in microbial communities that possess the metabolic ability to degrade xenobiotics, especially non-readily biodegradable chemicals (Kovárová-Kovar and Egli, 1998; Jung *et al.*, 2016). Second, the primers used to quantify the 2,4-DCP degraders might capture the broad range of putative 2,4-DCP degrading organisms. However, for many chemicals, there might be multiple pathways for complete transformation with involvement of multiple taxa (Copley, 2009). So, the functional gene based methods, used in this study, only capture and quantify the group of taxa contributing to specific steps in a particular pathway.

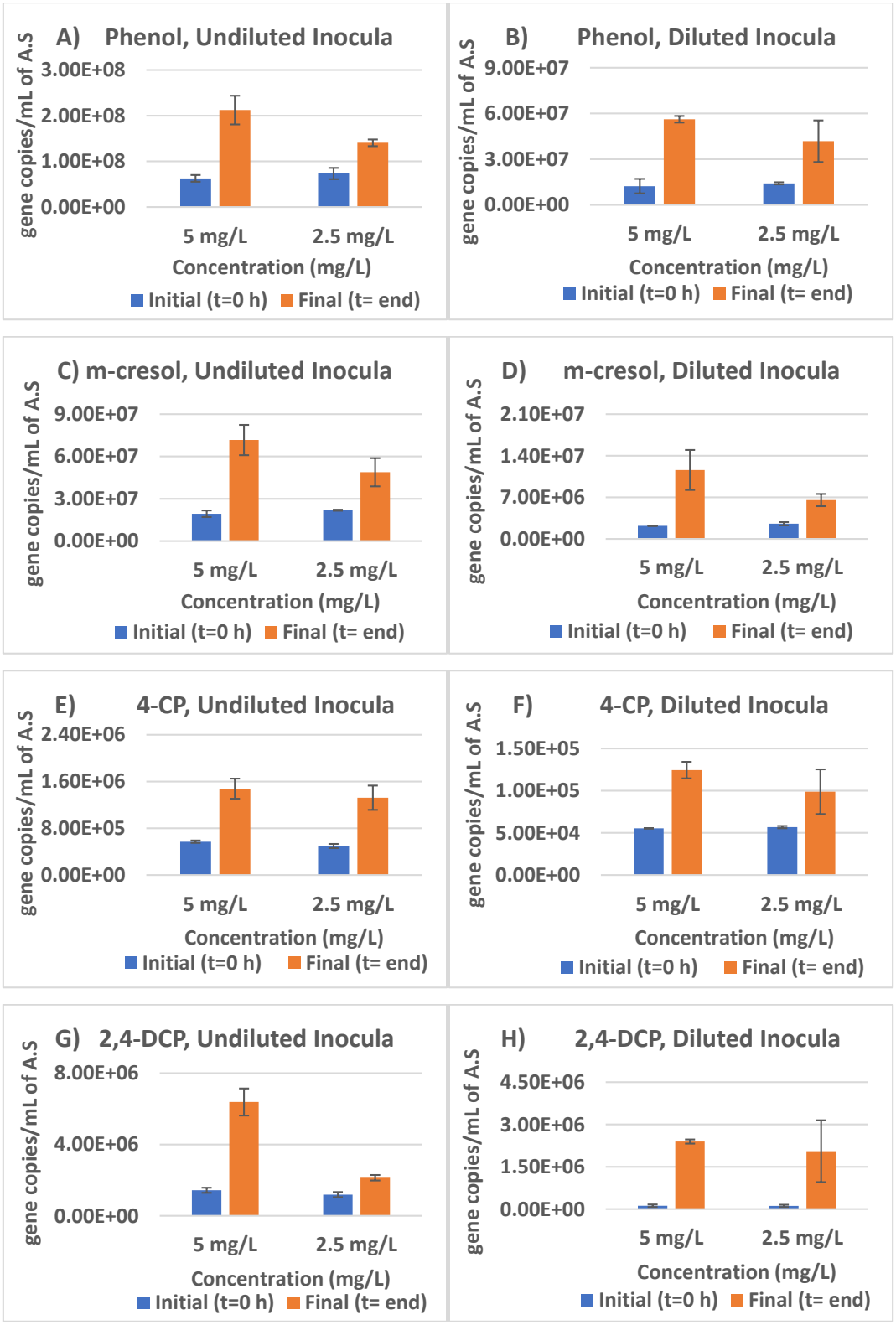


Figure 7-2 Major catabolic genes (at the beginning and end of biodegradation assays) involved in the degradation pathways of test chemicals for different experimental conditions; phenol (A and B), m-cresol (C and D), 4-chlorophenol (E and F) and 2,4-dichlorophenol (G and H). Two sample t-test was used to comapre; (i) the catabolic gene copy numbers at the beginning and end of the biodegradation assays, and (ii) final catabolic catabolic gene copy numbers in the inocula spiked with 2 different concentrations of chemicals. Data points are an average of biological duplicate samples (each biological sample was run in triplicate). Error bars are ± 1 s.d. of biological duplicate measurements.

Total cell count

The total bacterial cell counts were used to estimate the variability in the bacterial cell abundance in the activated sludge inocula in biodegradation assays following test chemical additions. There were no significant changes (2-sample t-test, p-values > 0.05) in the total bacterial counts at the end of the biodegradation assays with undiluted inocula at the two different spiked chemical concentrations (2.5 mg/L and 5 mg/L) (**Figure 7-3**) for all chemicals except 2,4-DCP (both concentration) and m-cresol (5 mg/L). For these three conditions, bacterial counts at the end of assays were significantly different from the initial counts (2-sample t-test, p-value > 0.05), being lower in the case of 2,4-DCP and higher in m-cresol.

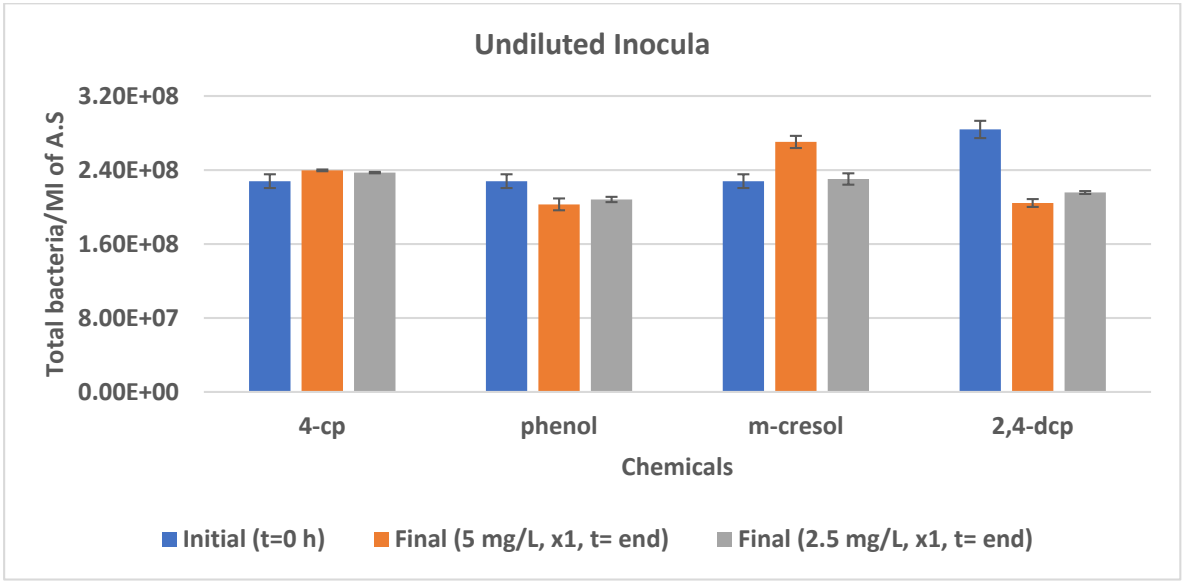


Figure 7-3 Comparison of total bacteria counts in reactors with undiluted inocula during the biodegradation assays of different test chemicals at two different spiked chemical concentrations. Data points are an average of biological duplicate samples (each biological sample was run in triplicate). Error bars are ± 1 s.d. of biological duplicate measurements.

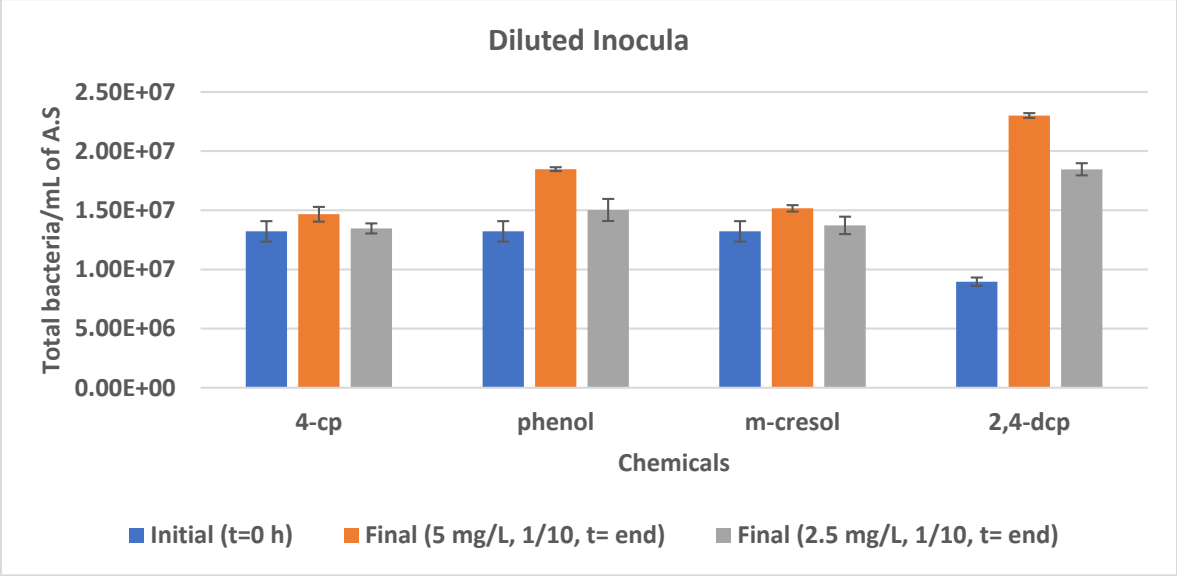


Figure 7-4 Comparison of total bacteria counts in reactors with diluted inocula during the biodegradation assays of different test chemicals at two different spiked chemical concentrations. Data points are an average of biological duplicate samples (each biological sample has technical triplicate). Error bars are ± 1 s.d. of duplicate measurements.

In biodegradation assays with diluted inocula, variable results were observed for all test chemicals at two different spiked chemical concentrations (2.5 mg/L and 5 mg/L) (**Figure 7-4**). The total bacterial counts at the end of the assays were significantly greater (2-sample t-test, p-value < 0.05) from the initial count for 2,4-DCP (both concentrations) and phenol (5 mg/L) assays. For the other chemicals and conditions, there were no significant differences between initial and final bacterial counts (2-sample t-test, p-value > 0.05). It has to be noted that in 2,4-DCP biodegradation assays, the final bacterial count increased significantly at the end of assays in reactors with diluted inocula for both spiked concentrations. In contrast, total bacterial counts decreased significantly in reactors with undiluted inocula. In general, the total cell count results indicate biodegradation of chemical in activated sludge might follow the Monod-like no growth kinetics as mentioned elsewhere (Simkins and Alexander, 1984). While, the catabolic genes copy number analysis suggests putative degraders get competitive advantage over the duration of biodegradation assay and biodegradation can be associated with growth linked kinetic. To evaluate this, the degradation data obtained from different experimental conditions were fitted to the Monod no growth model equation. The regression analysis showed that data fit well to Monod-like no growth kinetics (**Figure 7-5** and **Figure 7-6**), as indicated by high R^2 values (**Table 7-4** and **Table 7-5**). This indicates that, at the spiked chemical concentration, growth of putative chemical degraders could be proportionally insignificant due to the high initial bacterial biomass in the inocula. Nevertheless, variation in the initial putative degrader number has a significant impact on food to microorganisms ratio (F:M), which ultimately influence the degradation kinetics (Pitter and Chudoba, 1990;

McCarty, 2012; Kowalczyk *et al.*, 2015b). For instant, when there is high chemical concentration and low biomass number, the biodegradation can be potentially growth linked; as the chemical concentration decreases and biomass number increases, the kinetics of biodegradation will differ, typically with less growth. Therefore, there is a need of further modelling, and a more complete sampling regimen to cover biomass concentration at each point, in order to accurately describe the kinetics of biodegradation. The kinetics of chemical degradation in the activated sludge involving putative degraders are proposed and discussed in more details in **Section 7.6**.

Table 7-4 Summary table for Monod no – growth model fits for each chemical. R^2 value reported is relationship between the observed chemical concentration and the model predicted chemical concentration. H and L for C_0 refer to 5 mg/L and 2.5 mg/L respectively while L for X_0 refer diluted (10-fold) inocula. 4-CP and 2,4-DCP refer to 4-chlorophenol and 2,4-dichlorophenol respectively.

Chemical	m-cresol		Phenol		4-CP		2,4-DCP	
	C_0=H X_0=L	C_0=L X_0=L	C_0=H X_0=L	C_0=L X_0=L	C_0=H X_0=L	C_0=L X_0=L	C_0=H X_0=L	C_0=L X_0=L
R^2	0.99	0.89	0.99	0.98	0.965	0.897	0.977	0.949
Residual Std. Error	1.01	0.208	0.0671	0.091	0.3427	0.391	0.113	0.1531
Degrees of Freedom	5	5	5	5	6	6	5	5

Table 7-5 Summary table for Monod no – growth model fits for each chemical. R^2 value reported is relationship between the observed chemical concentration and the model predicted chemical concentration. H and L for C_0 refer to 5 mg/L and 2.5 mg/L respectively while H for X_0 refer undiluted inocula. 4-CP and 2,4-DCP refer to 4-chlorophenol and 2,4-dichlorophenol respectively.

Chemicals	m-cresol		Phenol		4-CP		2,4-DCP	
	C_0=H, X_0=H	C_0 = L, X_0 = H	C_0=H, X_0=H	C_0 = L, X_0 = H	C_0=H, X_0=H	C_0 = L, X_0 = H	C_0=H, X_0=H	C_0 = L, X_0 = H
R^2	0.95	0.975	0.987	0.98	0.995	0.9667	0.983	0.965
Residual Std. Error	0.36	0.125	0.1985	0.153	0.1074	0.1846	0.1205	0.135
Degrees of Freedom	6	6	5	5	6	5	5	5

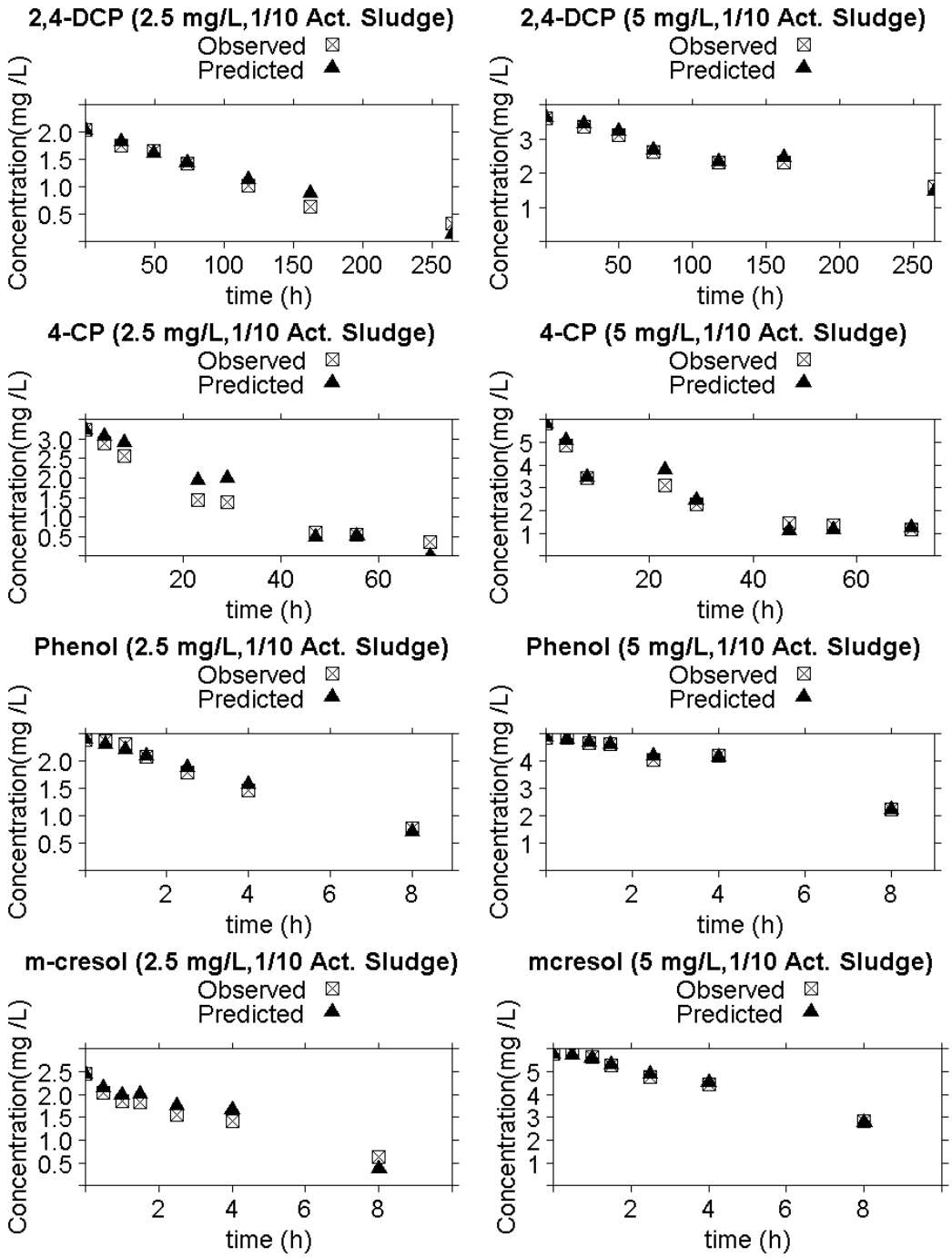


Figure 7-5 Monod no-growth curves, plotted for model predicted chemical concentration and the observed chemical concentration, for four chemicals (4-CP, 2,4-DCP, phenol and m-cresol) in diluted (10-fold) activated sludge inocula

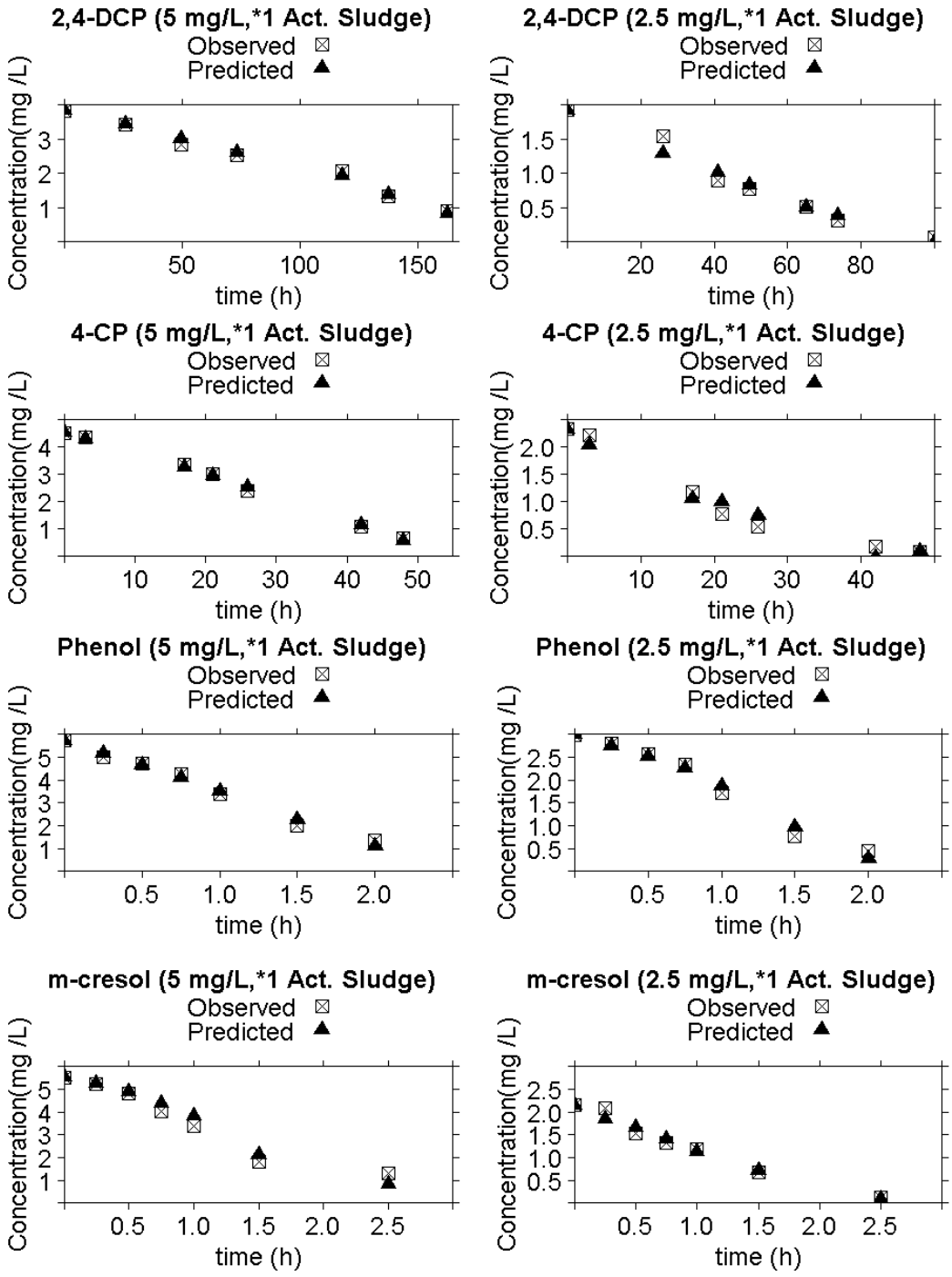


Figure 7-6 Monod no-growth curves, plotted for model predicted chemical concentration and the observed chemical concentration, for four chemicals (4-CP, 2,4-DCP, phenol and m-cresol) in undiluted activated sludge inocula

7.4.3 Microbial community structure

A mapping of the Bray-Cruris similarities of the OTUs at genus level (square root transformed) is presented in **Figure 7-7**. The result clearly showed that there was a 40 % similarity in microbial communities of all the samples under investigation regardless of the type of chemical dosed and dilution of inoculum used. The samples before the start of the biodegradation assays (except the t_0 inocula from 2,4-DCP degradation assays with undiluted inocula) and end samples from phenol, m-cresol and 4-chlorophenol (both inocula

concentration) assays clustered more closely and showed similarity of 50%. It has to be noted that, the initial inocula for 2,4-DCP and other three test chemicals biodegradation assays were sampled from WWTP4 on different days. The final (end of biodegradation assays) microbial communities in the 10-fold diluted inocula spiked with 2,4-DCP differed substantially from the original inoculum source. This might be attributed to the reduction in the microbial diversity caused by dilution and addition of 2,4-DCP, which is toxic to many microbes unable to metabolize 2,4-DCP in the inocula, and might have provided suitable conditions for the growth of rare taxa capable of degrading the test chemical.

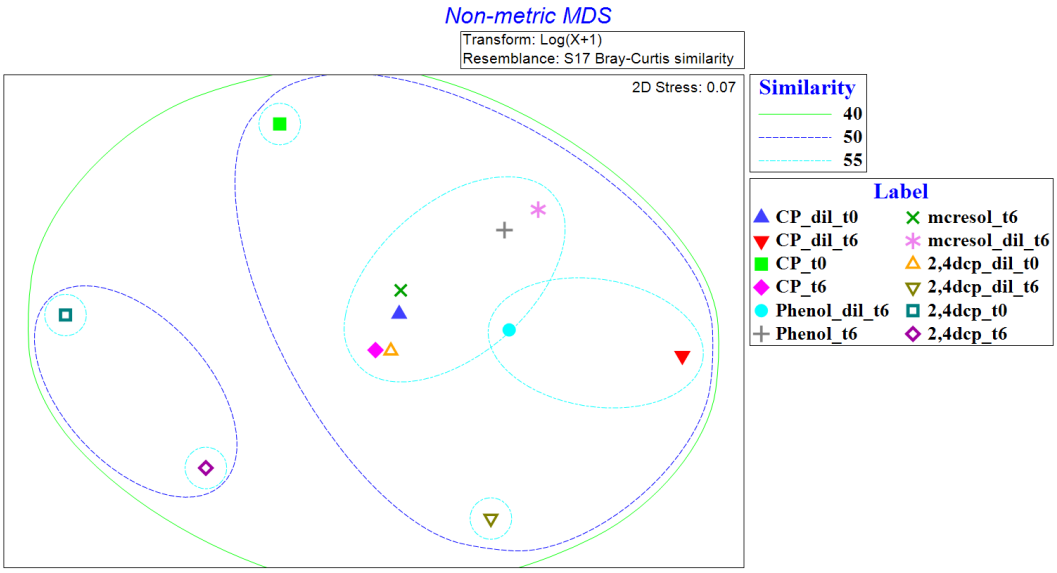


Figure 7-7 Multidimensional scale plot comparing the bacterial communities based on genus level. The communities were taxonomically characterized at the beginning (i.e. t0) and end (i.e.t6) of biodegradation assays

7.4.4 OTU functions in test chemical biodegradation

Test chemicals appeared to affect the dominance of certain families and genera. In all biodegradation assays spiked with 5 mg/L of test chemical, the relative abundance of different families and genera had significantly increased or decreased bythe end of biodegradation assays (**Figure 7-8** and **Figure 7-9**, and **Figure D 4 –D 8** [**Appendix D**]), suggesting a shift in microbial community structure following addition of aromatic chemicals to the inocula. The microbial community structure of different environment samples in microcosm studies has often been reported to shift following the addition of xenobiotics (Zhang *et al.*, 2008; Jung *et al.*, 2016; Mangse, 2016). This may be expected because the addition of xenobiotics changes the nutrient dynamics of the system. As a result, some microorganisms may simply diminish as they are starved of growth substrate, others may grow, and others toxically inhibited by the chemical, others may not do anything but persist. There was a significant difference with respect to the relative abundances of the OTUs between the inocula at time 0 and the inocula at the end of biodegradation assays for all the test chemicals (**Table 7-7** and

Table 7-8). There was a significant increase (G-test, $p < 0.05$) in the relative abundance of *Pseudomonas, Ralstonia, Dokdonella* and *Aequorivita* genera in 2,4-DCP spiked reactors with undiluted inocula. Similarly, in undiluted inocula, relative abundances of *Pseudomonas, Acinetobacter, Alcanivorax, Rhodanobacter, Dokdonella, Ralstonia, Fluviicola* and *Sphingopyxis* were significantly higher. 4-chlorophenol mostly promoted the abundances of *Pseudomonas, Rhodanobacter, Paracoccus* and *Rastolnia*, regardless of the dilution of initial inoculum. Similarly, in phenol-spiked inocula, *Rhadanobacter, Pseudomonas, Paracoccus, Ralstonia, Luteimonas, Sphingopyxis* and *Spingobacterium* relative abundances were significantly higher at the end of biodegradation assays (**Figure 7-10**). Likewise, m-cresol mostly promoted *Pseudoxanthomonas, Pseudomonas, Rhadanobacter, Paracoccus*, Luteimonas and *Ralstonia*. Increment in the relative abundance of the aforementioned genera also resulted in increase in their rank abundance. A ranking of OTUs at the genus level of taxonomy in the phenol degradation assays under studied conditions (both diluted and undiluted inocula) based on their absolute abundances to the total absolute abundance in the overall data set are shown in **Figure 7-11**. Those genera that were significantly different [i.e. initial (t_0) and final (t_{end})] were only included in the rank abundance study.

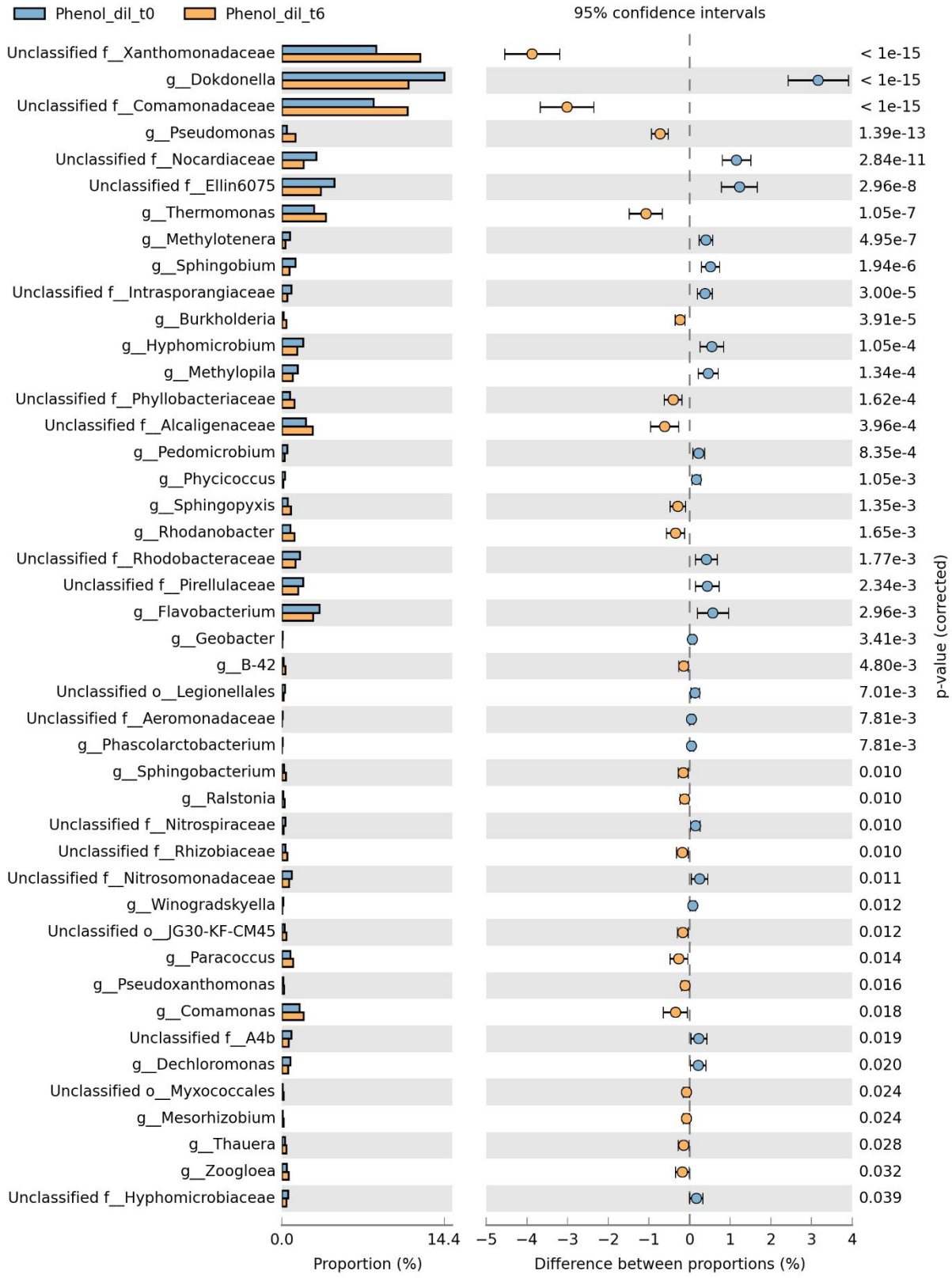


Figure 7-8 Extended error bar plot considering the abundance profile of microbial domains in the 16S rRNA amplicon sequencing data for phenol degradation assay in diluted inoculum. The bacterial domain that increased or decreased significantly (G-test, $p<0.05$) over the duration of assay are reported in the plot.

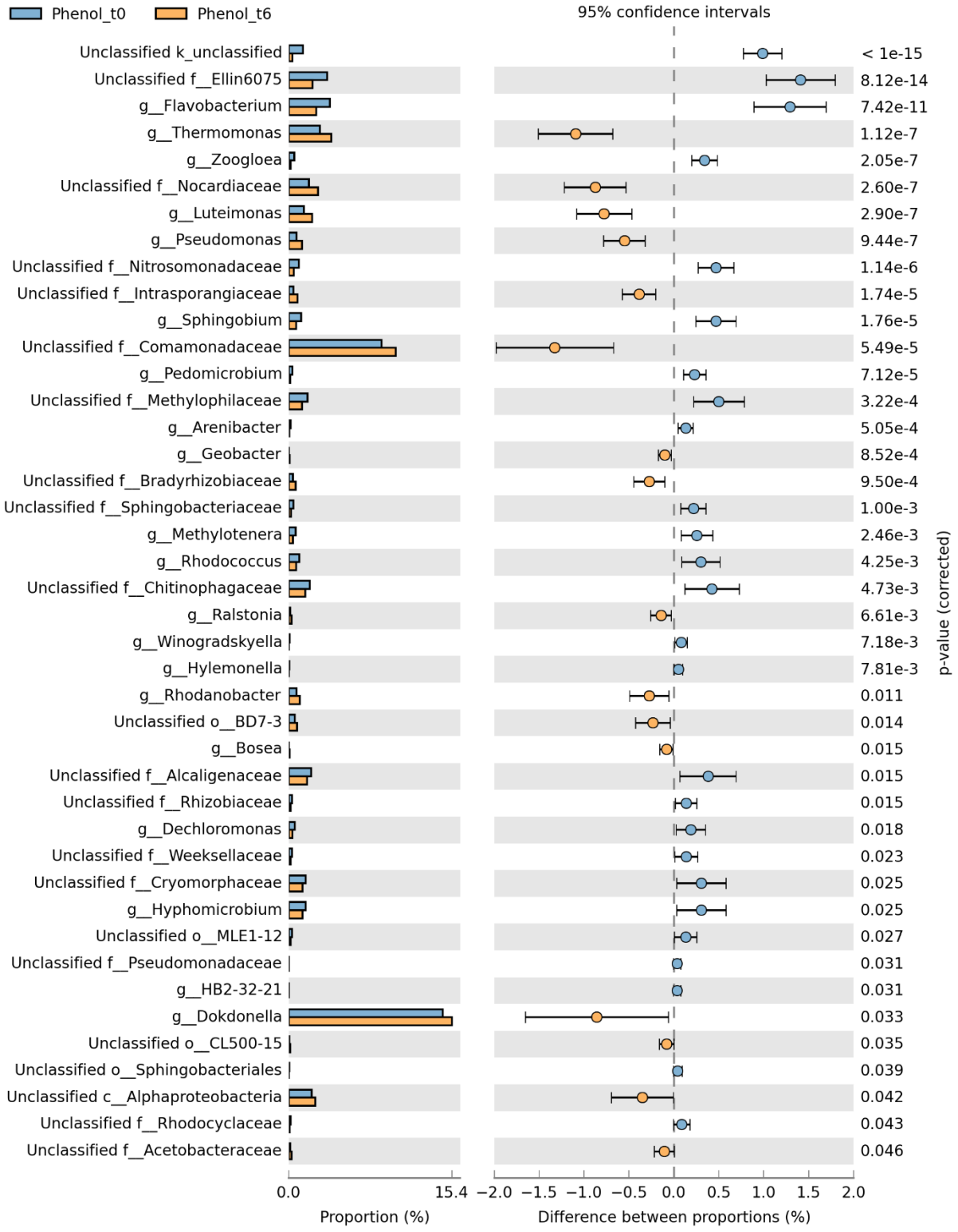


Figure 7-9 Extended error bar plot considering the abundance profile of microbial domains in the 16S rRNA amplicon sequencing data for phenol degradation assay in undiluted inoculum. The bacterial domain that increased or decreased significantly (G-test, $p<0.05$) over the duration of assay are reported in the plot.

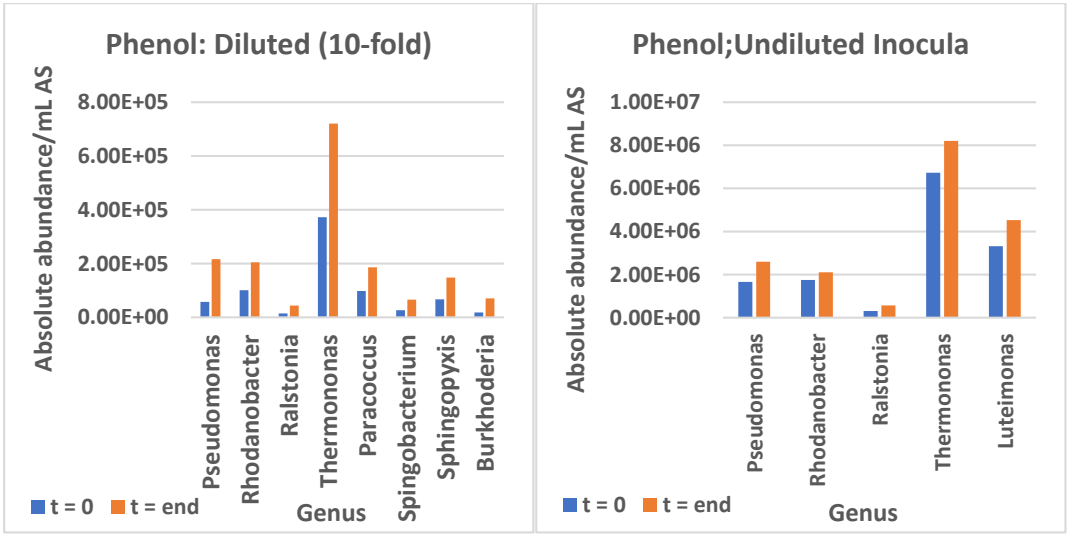
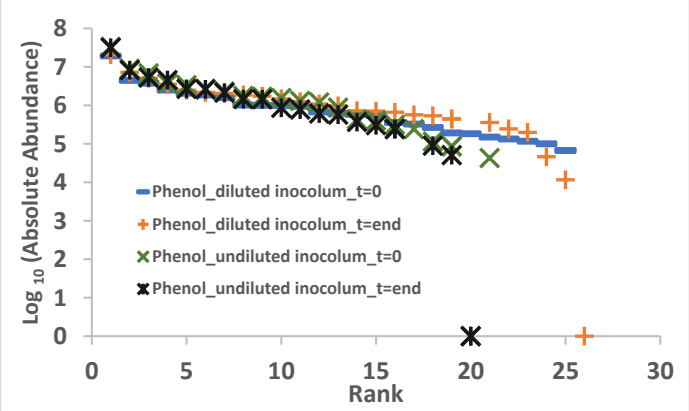


Figure 7-10 Comparison of initial (t = initial) and final (t = final) absolute abundance of putative phenol degraders in the inocula used for phenol degradation assay at two different inocula concentration. 5 mg/L of phenol was spiked in both conditions. The absolute abundance of each putative phenol degrader was estimated by multiplying the relative abundance of the degrader with the total bacteria count/mL of activated sludge estimated from flow cytometry.



	Rank (Undiluted Inocula)		Rank (Diluted Inocula)	
Genus	t = 0	t = end	t = 0	t = end
Pseudomonas	9	6	14	6
Rhodanobacter	8	7	8	7
Ralstonia	16	13	21	19
Thermomonas	3	2	3	2
Luteimonas	5	4		
Paracaoccus			9	8
Sphingobacterium			18	16
Sphingopyxis			12	10
Burkholderia			19	15

Figure 7-11 Left: Absolute abundance (Log transformed) of genera that significantly changed over the duration of phenol biodegradation assays (both conditions) in rank order. Right: The rank order of the putative phenol degraders at the start (t = 0) and end (t = end) of the assays.

Pseudomonas, *Burkholderia*, *Acinetobacter*, *Pseudoxanthomonas* and *Ralstonia* are Gram-negative, aerobic genera belonging to the families Pseudomonadaceae, Burkholderiaceae, Moraxellaceae, Xanthomonadaceae and Ralstoniaceae respectively. Taxa within these genera have been reported to metabolize a wide range of organic compounds (e.g., BTEX, phenolics, chlorinated phenolics and PAHs) (van Beilen *et al.*, 1994; Steinle *et al.*, 1998; Shaw *et al.*, 1999; Reardon *et al.*, 2000; Palleroni *et al.*, 2010; Mukherjee and Bordoloi, 2012; Choi *et al.*, 2013; Chen *et al.*, 2017). *Pseudomonas* is a metabolically versatile genus that comprises microorganisms which can live aerobically or anaerobically on nitrates as an electron acceptor (Palleroni *et al.*, 2010). The species belonging to this genus are reported to have the potential to metabolise a wide range of organic compounds ranging from alkanes (van Beilen *et al.*,

1994) to aromatic hydrocarbons – cresols, phenols chlorophenols and BTEX (Reardon *et al.*, 2000; Mukherjee and Bordoloi, 2012; El-Naas *et al.*, 2017) under aerobic conditions. Along with putative specific degraders, some genera containing taxa that are capable of degrading multiple polyaromatic hydrocarbons (PAHs), crude oil hydrocarbon and pharmaceuticals were also enriched over the duration of assays (**Table 7-7** and **Table 7-8**). This could be attributed to the availability of multiple nutrients and substrates within the activated sludge inocula, as the influent for WWTPs mostly constitute hormones, pharmaceuticals, petrochemicals, personal care products, and dyes (Benotti *et al.*, 2008; Grassi *et al.*, 2012). The results also indicated that there were significant decreases (G-test, $p < 0.05$) in the relative abundance of a significant number of genera and families **[Figure 7-8 and Figure 7-9, and Figure 18-23 (Appendix D)]** following exposure of inocula (both diluted and undiluted) to all four test chemicals. This most likely results from substrate depletion or toxicity exerted by the chemicals on the microbial population. It has to be noted that, the reactor used in my study are closed batch reactor system and have limited available substrate and space. As a result, there is always a competition among the microbial population for the available resources and space, the organisms that can resist the stress have competitive advantage. On the other hand, the spiked chemicals might also exert a toxic effect to some of the microbial population within the communities and ultimately results in decrease in their abundances within the community.

7.4.5 Distribution of rare and abundant taxa in microbial community

To investigate how the different taxa are distributed in the microbial communities of studied inocula and their abundance changes over the duration of time (i.e. biodegradation assays) with addition of test chemicals (phenol, m-cresol, 4-chlorophenol and 2,4-DCP), the log(x+1) –transformed value of percentage abundance of each taxa at time zero (start of biodegradation assay) was plotted versus percent abundance of each taxa at the end of assays (**Figure *7-12*** and **Figure 7-13**; for phenol biodegradation assays).

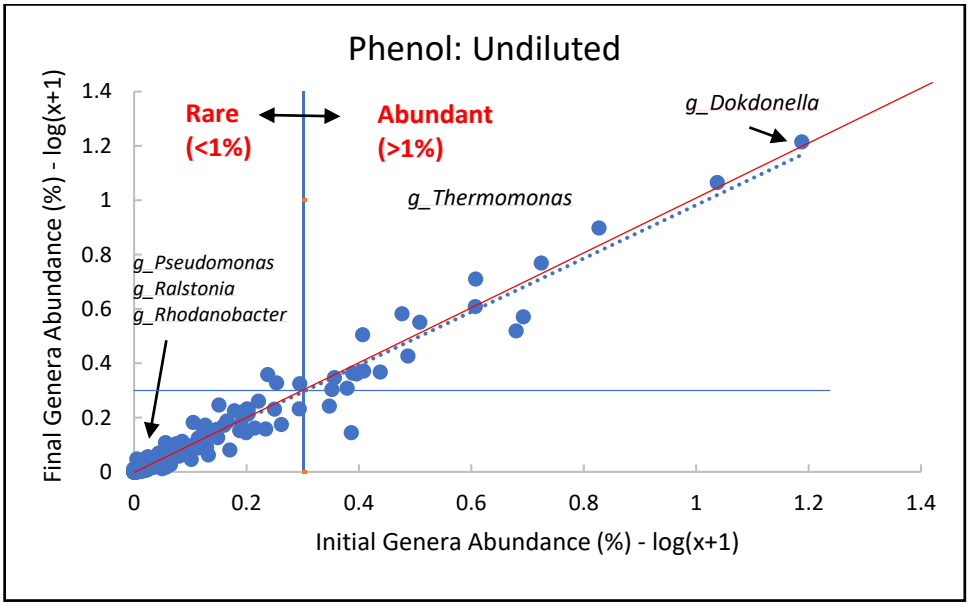
Vuono and colleagues (Vuono *et al.*, 2016) have suggested 1% abundance as an upper threshold of rarity in activated sludge communities. However, this threshold is arbitrary and subjective, and can increase or decrease depending on the microbial diversity of the environment concerned. Nevertheless, in this study, 1% abundance was considered as an upper threshold of rarity. Of the 601 genera present study-wide, on average 21 genera ± 0.95 were initially present in the inocula (both diluted and undiluted) with a relative abundance of ≥ 1%, while the remaining taxa were present with a relative abundance of <1%. For all the identified genera, their relative abundance either decreased or increased over the duration of

assay. The genera above the x=y line were enriched while those below represent genera whose abundance decreased at the end of biodegradation assays.

The majority of the aforementioned putative degraders (**Table 7-7** and **Table 7-8**) were present at a relative abundance of <1% and are thus relatively rare taxa, with few exceptions. For example, *Dokdonella* and *Thermomonas*, which are known to have metabolic potential to degrade poly-aromatic hydrocarbon (Mergaert *et al.*, 2003; Bacosa and Inoue, 2015), were one of the few abundant genera identified in the activated sludge microbial communities. More interestingly, some of the putative degraders, which were defined as rare genera in the starting inocula, got enriched over the duration of experiment and became abundant genera (**Table 7-7** and **Table 7-8**). However, none of the genera had relative percentage abundance greater than 2.71%. It should also be noted that there were significant decreases (G-test, $p < 0.05$) in the relative abundance of a significant number of genera and families (**Figure 7-8** [A & B] and **Figure 7-9** [A & B]).

The relationship between the relative abundance of the genera in the inocula at the start and end of the biodegradation assays were assessed by univariate regression analysis (**Table 7-6**). The intercept of the regression lines was set to zero. In short, reported R^2 values reveal a significant correlation between initial and final relative abundance of the genera (p-value < 0.05. The slope of the regression line, which indicates how close the final and initial abundance are, regardless of genera, are reported in **Table 7-6**. The slope for the 5 studied conditions were greater than or equal to 0.98, while for diluted inocula spiked with 4-CP and, both undiluted and diluted inocula spiked with 2,4-DCP were 0.936, 0.934 and 0.89, respectively. Slope and R^2 values can be used together to evaluate the stability of total bacteria abundance in a microbial system. The observed significant correlation (**high $R^2 \geq$ 0.86**, exception; 10-fold diluted inocula spiked with 2,4-DCP, $R^2 = 0.80$) and slope of regression line close to 1 for all the studied chemicals at both inocula concentration suggest that there is a balance in the increment and decrement of the relative abundance of the observed genera within the inocula. As a result, the total cell count in the inocula for those experimental conditions were maintained near constant value, over the duration of experiments (**Figure 7-3** and **Figure 7-4**). In 10-fold diluted inocula spiked with 2,4-DCP, the observed R^2 value and slope was comparatively lower than the R^2 values and slope for other experimental conditions, and in that system the final cell count of the inocula was twice as high as initial count.

[A]



[B]

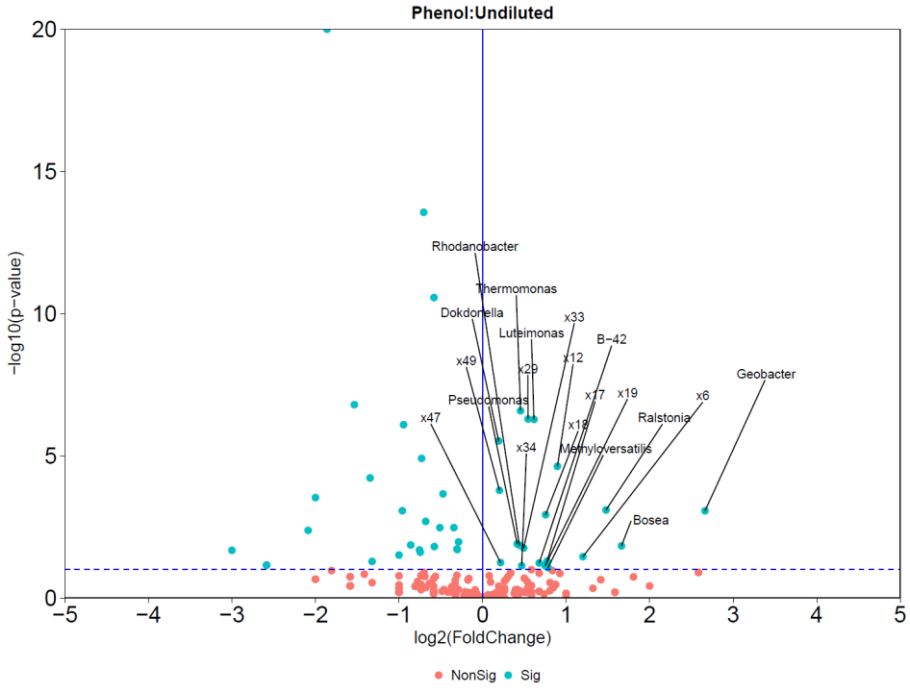
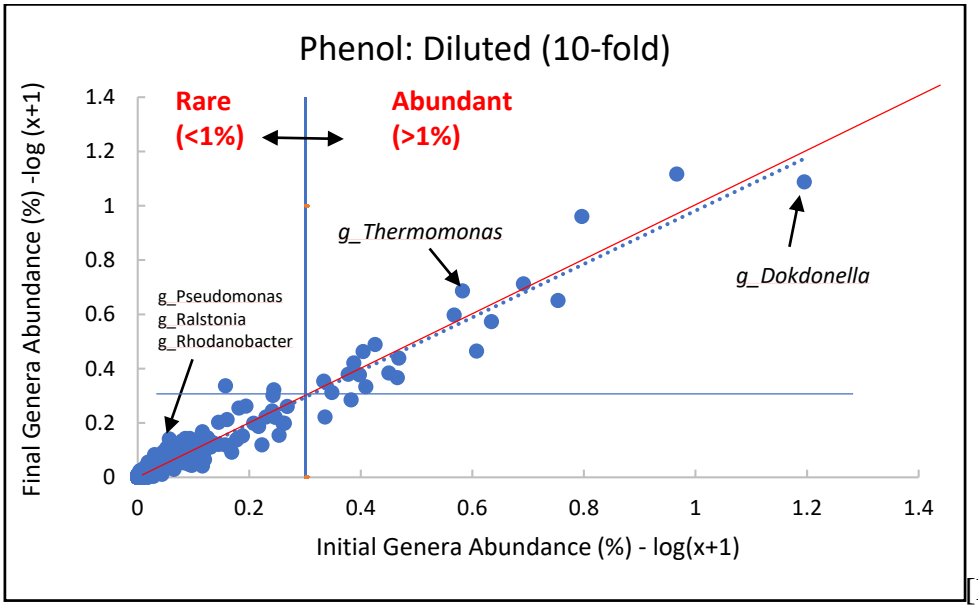


Figure 7-12 [A] Relationship between final and initial abundances for observed genera in phenol biodegradation assays with undiluted activated sludge inocula. The solid red line is x=y line andthe dotted blue line is the fitted trend line for the univariate regression analysis. The solid vertical and horizontal blue line delineates rare (<1% relative abundance) and abundant genera (>/= 1% relative abundance) in the inoculum initially and at the end of degradation assays, respectively. [B] Volcano plot showing the 16S amplicon sequencing data at genus level (L6). Black lines pointed towards the blue solid points indicate genera (both classified and unclassified [indicated by x_n, where n = 1,2,3,..n]) that enriched atleast 2-fold (x axis) over the duration of experiment and high statistical significance (-log10 p-value, y-axis). The dashed blue line shows where p-value = 0.05 with point above the line having p-value < 0.05 and points below the line having p > 0.05. The relative abundance of genera on the left side of solid blue line decreased while on the right got enriched over the duration of experiment.)

[A]



[B]

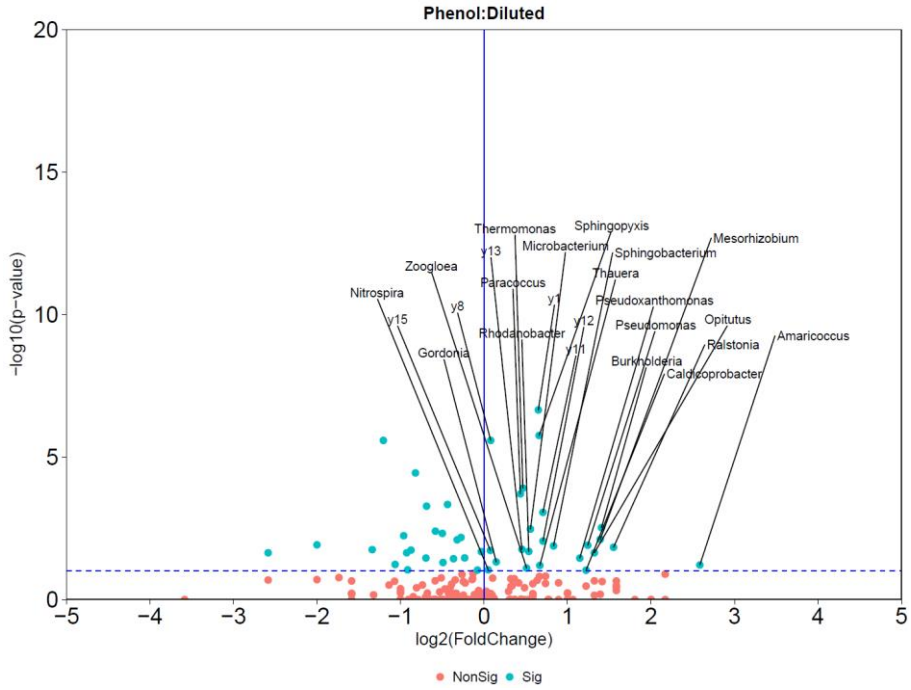


Figure 7-13 [A] Relationship between final and initial abundances for observed genera in phenol biodegradation assays with diluted (10-fold) activated sludge inocula. The solid red line is x=y line andthe dotted blue line is the fitted trend line for the univariate regression analysis. The solid vertical and horizontal blue line delineates rare (<1% relative abundance) and abundant genera (>/= 1% relative abundance) in the inoculum initially and at the end of degradation assays respectively. [B] Volcano plot showing the 16S amplicon sequencing data at genus level (L6). Black lines pointed towards the blue solid points indicate genera (both classified and unclassified [indicated by y_n, where n = 1,2,3, .. n]) that enriched atleast 2-fold (x axis) over the duration of experiment and high statistical significance (-log10 p-value, y-axis). The dashed blue line shows where p-value = 0.05 with point above the line having p-value < 0.05 and points below the line having p > 0.05. The relative abundance of genera on the left side of solid blue line decreased while on the right got enriched over the duration of experiment.

Table 7-6 Summary statistics of univariate regression analysis between final and initial relative abundance of observed genera in the degradation assays performed for four different test chemicals (5 mg/L) at two different inocula concentration.

	C_0 = 5 mg/L and X_0 = Undiluted				C_0 = 5 mg/L and X_0 = diluted (10-fold)			
Chemical	Phenol	m-cresol	4-CP	2,4-DCP	Phenol	m-cresol	4-CP	2,4-DCP
Slope	0.982	0.98	0.98	0.934	0.981	0.98	0.936	0.89
R^2	0.964	0.978	0.951	0.902	0.958	0.950	0.861	0.806
SE	0.021	0.017	0.025	0.034	0.023	0.026	0.043	0.051
DF	600	600	600	600	600	600	600	600
p-value	0	0	0	0	0	0	0	0

These results highlight the importance of the rare species pool in the microbial community, which are responsible for the degradation of environmentally or economically important chemicals. These results are in line with the previous study conducted by Vuono et al (2016) (Vuono *et al.*, 2016), where they showed that the activity of rare taxa were significantly higher than abundant taxa and played a significant role in chemical degradation. Furthermore, results also indicate that, microbial degradation might bring about the enrichment of putative chemical degraders, which not necessarily influence the overall cell count of the system.

7.5 Implication of 16S rRNA amplicon sequencing and functional gene analyses: rare taxa identification and screening of inocula

The genera that might harbour putative chemical degraders were identified in the activated sludge inocula in the previous section **(Section 7.4.4).** However, it can be ignored the fact that selecting the genera that could harbour putative degraders is speculative, and thus demands for a use of confirmative techniques, for example Stable Isotope Probing technique, which can identify the actual degraders responsible for the degradation of chemical. Nevertheless, the relative abundance of those genera was multiplied by the total cell count (bacteria/mL of A.S), in order to estimate the absolute abundance of putative chemical degraders. Whereas, the major catabolic genes involved in the test chemicals biodegradation were quantified and presented in **Section 7.4.2.** It was assumed that each organism in the microbial community contained one copy of the gene. Finally the comparison between the absolute abundance of putative degraders estimated with 16S rRNA sequencing and those quantified from functional gene analysis was performed (**Figure 7-14** and **Figure 7-15**).

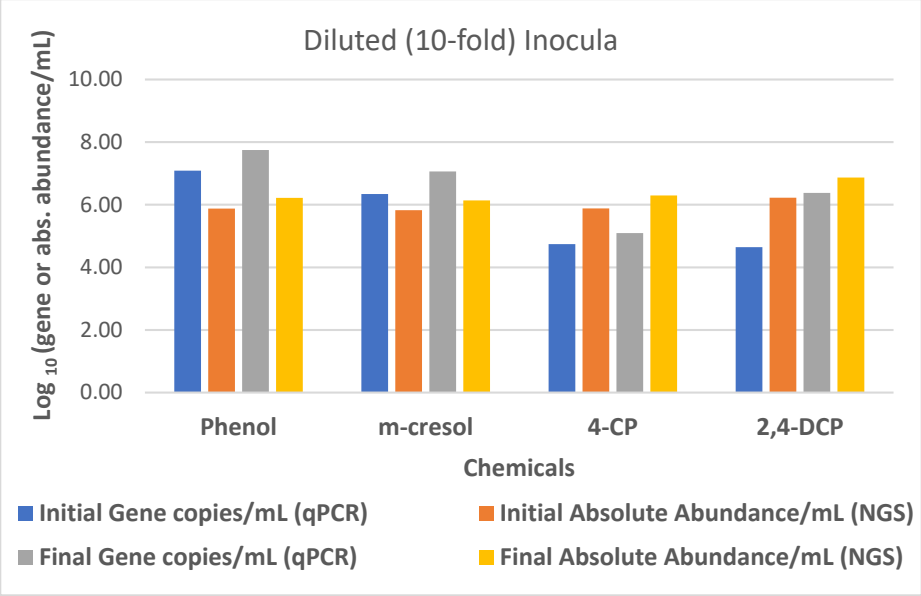


Figure 7-14 Comparison between initial (t_0) and final (t_{final}) absolute abundance of putative chemical degraders estimated with functional gene (with qPCR) and 16S rRNA amplicon sequencing (i.e. NGS) approach in biodegradation assays conducted with diluted inocula.

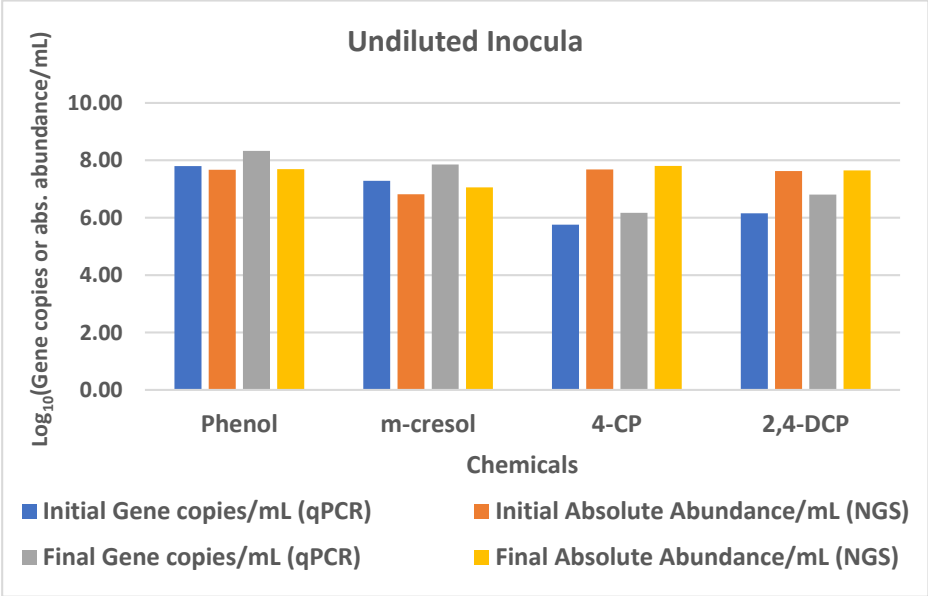


Figure 7-15 Comparison between initial (t_0) and final (t_{end}) probable absolute abundance of putative chemical degraders estimated with functional gene (with qPCR) and 16S rRNA amplicon sequencing approach (i.e. NGS) in biodegradation assays conducted with undiluted inocula.

In general, regardless of time point and inocula concentration, phenol and m-cresol putative degraders estimated with functional gene analysis were comparatively higher than those estimated from the 16S rRNA amplicon sequencing method. In contrast, the opposite was observed for 4-chlorophenol and 2,4-DCP putative degraders. These results suggest different possibilities. First, it is most likely that there are more putative degrader genera within the microbial community that still need to be discovered. In this study, the identified putative chemical degraders were previously known to degrade the chemicals under investigation. Secondly, it is possible that multiple copies of a gene might be present in some bacterial

species, which will ultimately influence the abundance of taxa observed by functional gene qPCR-based method (Pérez-Pantoja *et al.*, 2010). *Perez-Pantoja et a*l highlighted that the metabolically versatile bacteria can possess two or more copies of the genes that help them to adapt under changing environmental circumstances. For example, in *Cuprividus nectar* JMP134, there are two copies of phenol hydroxylase (PH), where one copy is associated with a catechol-1,2- dioxygenase, and other with a catechol-2,3-dioxygenase (Pérez-Pantoja *et al.*, 2010). It is important to note that, both approaches have their own limitations. The functional gene method only captures the group of taxa contributing to specific steps in the degradation pathway (Copley, 2009; Kowalczyk *et al.*, 2015a). A chemical can be degraded by a number of different putative degraders via multiple pathways with the involvement of different putative genes. For examples, phenol can be degraded by numerous organisms using multiple pathways (Gao *et al.*, 2010). The biological marker used in this study can only detect and quantify a group of phenol degraders that perform the phenol degradation via catechol pathway, while it is unable to detect degraders that used other genes (for example, 4-hydroxy-decarboxylase or phenyl phosphate synthase) for phenol degradation (Gao *et al.*, 2010). The *in situ* identification of actual degraders responsible for the degradation of chemical can be achieved by using stable isotope probing (SIP) that enables the identification of degraders that metabolize enriched chemicals (Wackett, 2004; Kowalczyk *et al.*, 2015a). However, developing the SIP method to identify the active degraders for enriched chemicals was outside the scope of this study. On the other hand, the result of the 16S rRNA amplicon sequencing is highly influenced by the library preparation methods and the choice of primers (Schirmer *et al.*, 2015; Tan *et al.*, 2015). In addition, the short sequence read length further compromise the taxonomic resolution and ultimately hinders better identification of putative taxa. Furthermore, downstream data processing like, methods used in chimera removal (Kennedy *et al.*, 2014) and OTU clustering (e.g., UCLAST) (Tan *et al.*, 2015) can also produce biases and ultimately have an influence in determining the relative abundance and diversity of taxa in microbial communities. Despite the concerns mentioned above, a number of studies have demonstrated that 16S rRNA amplicon sequencing might offer a good approximation of the microbial species composition and relative abundance in samples (Ong *et al.*, 2013; Vignola *et al.*, 2017), however the accuracy depends on the nature of sample analysed and is likely to be influenced by factors discussed above (i.e. methods used for library preparation and primer choice). Nevertheless, if particular taxa are associated with the catabolic genes involved in a given chemical's degradation, Next Generation Sequencing (NGS) might represent a cost effective means of screening inocula, though it requires further validation and investigation.

Table 7-7 Summary of statistically significant treatment effects [time zero undiluted activated sludge inocula (i.e.before spiking the test chemicals) versus final inocula (i.e. end of biodegradation assays), G-test, p<0.05] for OTU identified at genus level, where member of genus reportedly degrade test chemicals and aromatic hydrocarbons. The bold values indicate the percentage abundance while un-bold represent absolute abundance.

Percentage relative abundance of diluted inocula										
	Phenol		**m-cresol**		**4-CP**		**2,4-DCP**		**Degradation potential**	**References**
Genus	**Initial**	**Final**	**Initial**	**Final**	**Initial**	**Final**	**Initial**	**Final**		
Acinetobacter							**0.0375** 3.36 x 10^3	**2.7125** 6.24 x 10^5	BTEX, chlorophenol ,phenol, PAHs	(Hao *et al.*, 2002; Lee *et al.*, 2011; Paisio *et al.*, 2014; Zhou *et al.*, 2016)
Pseudomonas	**0.437** 5.7x10^4	**1.168** 2.16x10^5	**0.0437** 5.77 x 10^4	**0.987** 1.45 x 10^6	**0.437** 5.77 x 10^4	**1.01** 1.48 x 10^5	**0.1** 8.96 x 10^3	**0.73** 1.68 x 10^5	BTEX, chlorophenol s, phenol, cresols, alkanes, PAHs	(Reardon *et al.*, 2000; Palleroni *et al.*, 2010; Mukherjee and Bordoloi, 2012; El-Naas *et al.*, 2017)
Rhodanobacter	**0.7625** 1. x 10^5	**1.1063** 2.05 x 10^5	**0.7625** 1.01 x 10^5	**1.2** 1.76 x 10^5	**0.7625** 1.01 x 10^5	**1.2313** 1.81 x 10^5	**0.6813** 6.1 x 10^4	**1.5813** 3.34 x 10^5	PAHs, 2,4-D	(Kanaly *et al.*, 2002)
Ralstonia	**0.1125** 1.49x10^4	**0.237** 4.38 x 10^4	**0.1125** 1.49 x 10^4	**0.218** 3.2 x 10^4	**0.1125** 1.49 x 10^4	**0.18** 2.65 x 10^4	**0.068** 6.09 x 10^3	**0.193** 4.44 x 10^4	BTEX, chlorophenol s, phenol,	(Steinle *et al.*, 1998; Plumeier *et*

									cresols, PAHs	*al.*, 2002; Santisi *et al.*, 2015; Ghosal *et al.*, 2016)
Alcanivorax							0	**0.125** 2.88 x $10^{\wedge 4}$	Crude oil	(Santisi *et al.*, 2015)
Dokdonella							**14.825** 1.33 x $10^{\wedge 6}$	**16.675** 3.84 x $10^{\wedge 6}$	PAHs	(Bacosa and Inoue, 2015)
Thermononas	**2.8188** 3.72x$10^{\wedge 5}$	**3.8938** 7.2 x $10^{\wedge 5}$	**2.8187** 3.72 x $10^{\wedge 5}$	**4.7** 6.91 x $10^{\wedge 5}$	**2.8187** 3.72 x $10^{\wedge 5}$	**6.6438** 9.77x $10^{\wedge 5}$	**2.6625** 2.39 x $10^{\wedge 5}$	**9.775** 2.25 x $10^{\wedge 6}$	poly(e-caprolactone)	(Mergaert *et al.*, 2003)
Aequorivita					**0.1625** 2.15 x $10^{\wedge 4}$	**0.4313** 6.34 x $10^{\wedge 4}$			PAHs	(Wang *et al.*, 2016)
Paracoccus	**0.7438** 9.82x$10^{\wedge 4}$	**1.0063** 1.86 x $10^{\wedge 5}$	**0.7437** 9.82 x $10^{\wedge 4}$	**0.9812** 1.44 x $10^{\wedge 5}$	**0.7437** 9.82 x $10^{\wedge 4}$	**1.0813** 1.59 x $10^{\wedge 5}$			BTEX	(Gusmão *et al.*, 2007)
Spingobacterium	**0.2** 2.64x$10^{\wedge 4}$	**0.3563** 6.59 x $10^{\wedge 4}$			**0.2** 2.64 x $10^{\wedge 4}$	**1.7375** 2.55 x $10^{\wedge 5}$			Estrogen, dye, herbicide	(Min, 2004; Haiyan *et al.*, 2007; Tamboli *et al.*, 2010)
Pseudoxanthomonas			**0.0875** 1.14 x $10^{\wedge 4}$	**0.1875** 2.76 x $10^{\wedge 4}$					BTEX, PAHs	(Lee *et al.*, 2008; Lee *et al.*, 2012; Choi *et al.*, 2013)

Sphingopyxis	**0.5063** 6.68x10^4	**0.8** 1.48 x 10^5	**0.5062** 6.68 x 10^4	**1.1125** 1.64 x 10^5	**0.5062** 6.68 x 10^4	**1.1** 1.62 x 10^5	**0.269** 2.41 x 10^4	**0.4562** (1.05 x 10^5)	2,4,6-trichlorophenol	(Aranda *et al.*, 2003)
Burkhoderia	**0.14** 1.85x10^4	**0.38** 7.03 x 10^4							BTEX, chlorophenols, phenol, cresols, PAHs	(Mars *et al.*, 1996; Shaw *et al.*, 1999; Revathy *et al.*, 2015; Chen *et al.*, 2017)

Table 7-8 Summary of statistically significant treatment effects [time zero diluted activated sludge inocula (i.e.before spiking the test chemicals) versus final inocula (i.e. end of biodegradation assays), G-test, p<0.05] for OTU identified at genus level, where member of genus reportedly degrade test chemicals and aromatic hydrocarbons. The bold values indicate the percentage abundance while un-bold represent absolute abundance.

Percentage relative abundance of undiluted inocula										
	Phenol		**m-cresol**		**4-CP**		**2,4-DCP**		**Degradation potential**	**References**
Genus	**Initial**	**Final**	**Initial**	**Final**	**Initial**	**Final**	**Initial**	**Final**		
Pseudomonas	**0.73** 1.66x10^6	**1.28** 2.6x10^6	**0.73** 1.66x10^6	**1.3** 3.51x10^6	**0.73** 1.66x10^6	**1.31** 3.14x10^6	**0.26** 7.38x10^5	**0.868** 1.77x10^6	BTEX, chlorophenols, phenol, cresols, alkanes, PAHs	(Reardon *et al.*, 2000; Palleroni *et al.*, 2010; Mukherjee and Bordoloi, 2012; El-Naas *et al.*, 2017)
Rhodanobacter	**0.768** 1.66x10^6	**1.040** 1.66x10^6			**0.768** 1.66x10^6	**1.131** 1.66x10^6			PAHs, 2,4-D	(Kanaly *et al.*, 2002)
Ralstonia	**0.137**	**0.281**	**0.137**	**0.33**	**0.137**	**0.25**	**0.118**	**0.225**	BTEX,	(Steinle *et al.*, 1998;

	3.12x10^5	5.7x10^5	3.12x10^5	8.91x10^5	3.12x10^5	5.99x10^5	3.35x10^5	4.59x10^5	chlorophenols, phenol, cresols, PAHs	Plumeier *et al.*, 2002; Santisi *et al.*, 2015; Ghosal *et al.*, 2016)
Dokdonella	**14.437** 3.29x10^7	**15.531** 3.15x10^7			**14.556** 3.32x10^7	**17.375** 4.16x10^7	**14.387** 4.09x10^7	**20.275** 4.14x10^7	PAHs	(Bacosa and Inoue, 2015)
Thermononas	**2.95** 6.73x10^6	**4.043** 8.21x10^6			**2.95** 6.73x10^6	**3.487** 8.36x10^6			poly(e-caprolactone)	(Mergaert *et al.*, 2003)
Aequorivita							**0.187** 5.31x10^5	**0.406** 8.28x10^5	PAHs	(Wang *et al.*, 2016)
Paracoccus			**0.556** 1.27x10^6	**0.75** 2.03x10^6	**0.556** 1.27x10^6	**0.812** 1.95x10^6			BTEX	(Gusmão *et al.*, 2007)
Luteimonas	**1.456** 3.32x10^6	**2.231** 4.54x10^6	**1.456** 3.32x10^6	**1.818** 4.91x10^6	**1.456** 3.32x10^6	**2.1** 5.03x10^6			PAHs	(Bacosa and Inoue, 2015)

7.6 Implication of quantifying key catabolic genes in predicting the biodegradation kinetics

In the environment, the bacteria with an ability to degrade a particular pollutant are a small fraction of total microbial cells, and do not represent a constant fraction of the population (Vuono *et al.*, 2016). In addition, the degree of enzyme systems induced during the degradation of pollutant varies and their quantification is laborious. Therefore, determination of biodegradation kinetics in such an undefined mixed culture (e.g., activated sludge) system is the equivalent to a "black box approach". As a consequence, the degradation kinetics parameters are assigned to total bacterial biomass, not to the competent degraders responsible for biological activity.

A group of key catabolic genes (enzymes) encoded in the DNA (chromosome or plasmid) of competent degraders catalyse the degradation of such pollutants in the environment. Therefore, the ability to detect and quantify such genes helps in estimation of key competent degrader's number in the environment, which should ultimately form the basis towards estimation of kinetic parameters with active degraders, not the total biomass. Based on the observed significant linear correlation between the apparent first-order rate of chemical removal and the ratio of key catabolic genes and chemical concentration (X_0/C_0) initially present in the environment, a differential equation (i.e. Equation 1) can be formulated to describe the substrate removal rate in the environment. This is the general expression of the chemical disappearance in a system in which only competent degrader densities and substrate concentrations determines the kinetic of degradation.

$$\frac{-dS}{dt} = k\,X_0\,(\frac{C}{C_0}) \quad (1)$$

Where C is the concentration of the chemical, C_0 is the initial chemical concentration, X_0 is the initial concentration of key catabolic genes and k is a rate constant for removal of chemical. However, the relationship only applies for the range of initial chemical concentration tested. It is assumed that concentration of the chemical decreases linearly until the chemical is exhausted. Thus, the overall chemical removal rate is a function of the ratio of the remaining amount of chemical to the amount initially present.

Integrating Equation 1 gives;

$$C = C_0 * e^{-k*t*\frac{X_0}{C_0}} \quad (2)$$

By substituting k_1 for $k * \frac{X_0}{C_0}$, equation 2 becomes

$$C = C_0 * e^{-t*k_1} \quad (3)$$

$$lnC = lnC_0 - t * k_1 \quad (4)$$

Equation 4 is a straight line equation with coordinate y = lnC and x= t and with slope k_1 and represents the mathematical model of substrate removal in the batch reactor system with inoculum possessing competent degraders for a given initial degrader and chemical concentration.

7.7 Conclusion

Collectively, these results highlight that chemical structure, initial chemical concentration and inocula concentration (and characteristics) are the key drivers of chemical biodegradation in the environment. The biodegradation rate of a chemical was shown to be a function of both initial chemical concentration and degrader number in an environment due to the strong and significant linear correlation between rate and ratio of X_0 and C_0 (i.e. X_0/C_0). The majority of key degraders of test chemicals are present at low abundance under natural conditions and might gain competitive advantages and thrive well when the initial biodiversity is impaired with chemical contamination. Furthermore, results also highlighted the importance of rare species pool in the microbial community, which are responsible for the degradation of chemicals in the activated sludge systems. In addition, this study also indicated that, microbial degradation might bring about the enrichment of putative chemical degraders, which not necessarily influence the overall cell count of the system. The combination of functional gene and NGS analyses performed in this study enabled inocula screening, and provided an insights into identifying and quantifying probable putative chemical degraders in the environment.

7.8 Verification and calibration of QSBR model

In the environment, the rate of chemical biodegradation is uncertain and difficult to reproduce as it is influenced by spatial and temporal variability of combination of different factors (i.e. abundance and activity of degrading microorganism, environmental conditions, structure and concentration of chemicals). On the other hand, the majority of existing QSBR, including that in Chapter 3, are developed with data from different biodegradation databases, which are often based on another form of QSBR. The chemical biodegradation data used in the development of such databases are derived from first tier biodegradation screening tests (typically OECD 301 or 302 tests) and rarely represent the true rates. The reliability of the predicted chemical degradation rate (at different environmental compartments) with existing QSBR models is a major concern (Peijnenburg and Damborský, 2012). This demands a method/approach that can be used to validate and/or calibrate the existing models, so that it is possible to apply the existing QSBR models to predict the chemical degradation rate in any environmental compartment.

Table 7-9 shows the summary statistics for the univariate regression analysis for four studied conditions. A strong and significant linear correlation ($0.74 \leq r^2 \leq 0.86$, p-value ≤ 0.1), along with similar patterns of biodegradation rates, were observed between the experimentally determined rates (different chemical and inocula concentrations) [**Figure 7-16**] and those rates used in the previously developed QSBR model (Chapter 3) for four of the test chemicals. Similar correlation was observed between rates determined with pure degrader and the rates used in QSBR models (**Chapter 4, Section 4.4.4 or Table 7-9**). The biodegradation rates of each chemical for the different studied conditions can be significantly different (one-way ANOVA, $p < 0.05$; **Table 7-2**) and thus suggest that, calibration of existing QSBR models is a pre-requisite before applying them to predict the chemical persistence in any environmental compartment. Obtaining the experimental rates (i.e. performing biodegradation assay with inocula from relevant environment) for several chemicals included in the dataset for QSBR model development (e.g. Chapter 3), and performing the univariate regression analysis, as done in this study might form a rational basis for calibrating these QSBR models with real rates.

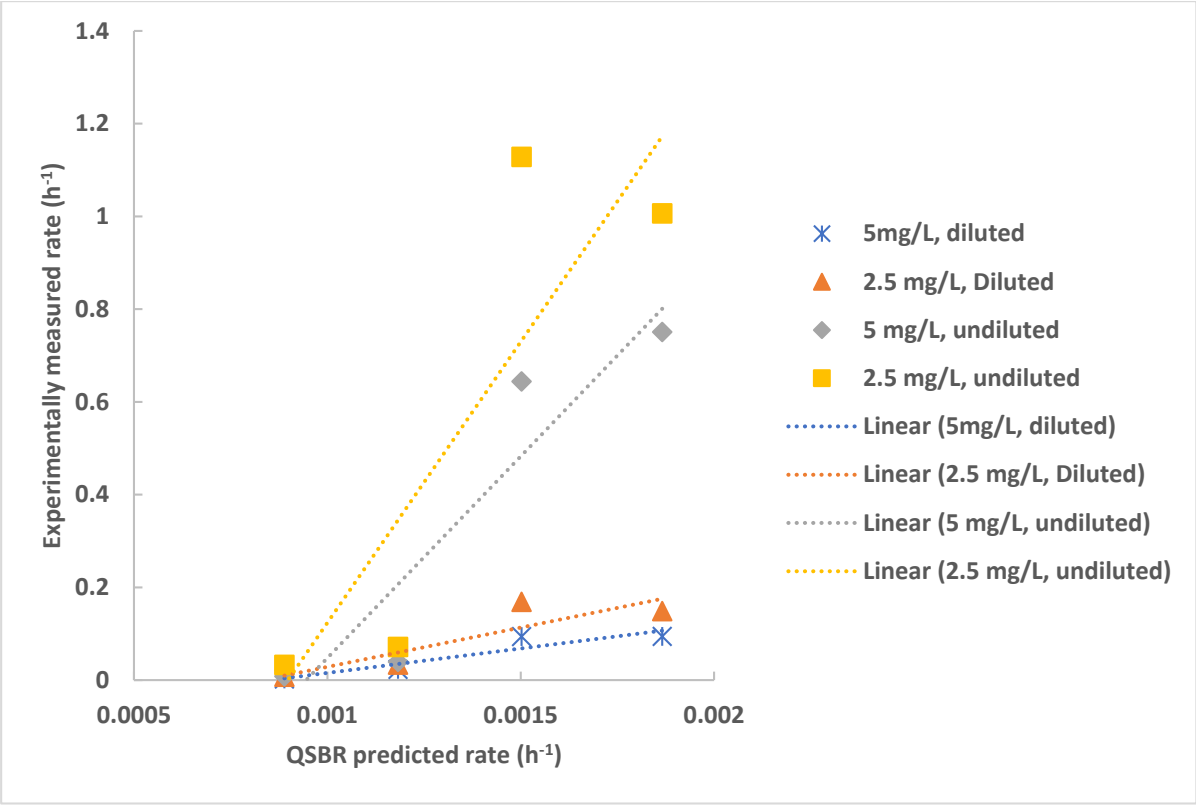


Figure 7-16 Relationship between the experimentally determined biodegradation rates in grab activated sludge inocula and the biodegradation rates used in previously developed QSBR models for four test chemicals. The figure legend indicates the reactor conditions (i.e. concentration of spiked chemicals and inocula concentration). Diluted refers to 10 fold dilution, whereas undiluted refers to raw activated sludge

Table 7-9 Summary statistics of univariate regression analysis between experimentally determined biodegradation rates at different studied conditions including pure culture experiments and the biodegradation rates used in previously developed QSBR models for four test chemicals. The high C_0 (initial spiked chemical concentration) indicates 5 mg/L while X_0 (initial putative chemical degrader numbers) indicates undiluted activated sludge, low C_0 indicates 2.5 mg/L while low X_0 indicates 10-fold diluted sludge with OECD mineral medium.

Experimental condition	**C_0 = high, X_0 = low**	**C_0 = low, X_0 = low**	**C_0 = high, X_0 = high**	**C_0 = low, X_0 = high**	**Pure culture experiment (Chapter 4)**
R^2	0.86	0.774	0.87	0.748	0.85
Residual Std. Error	0.021	0.047	0.17	0.361	0.0002
Degrees of Freedom	3	3	3	3	4
P-value	0.07	0.1	0.066	0.13	0.023

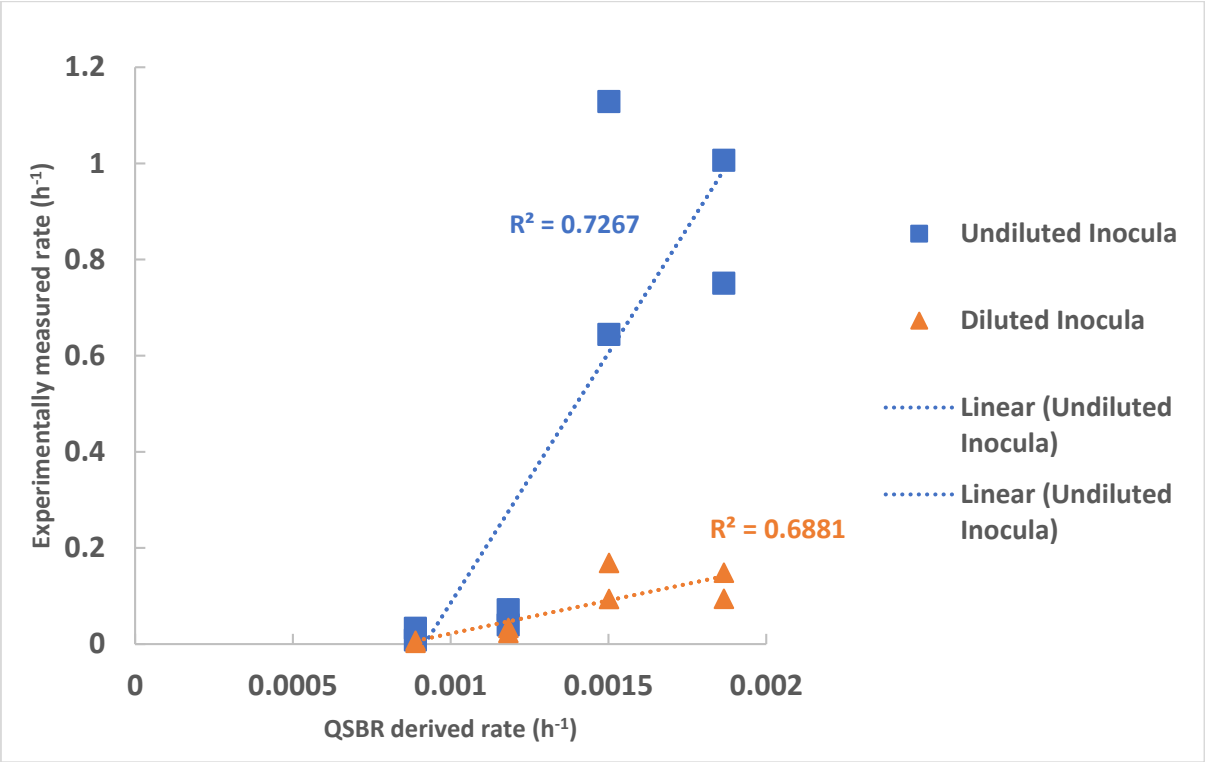


Figure 7-17 Relationship between the experimentally determined biodegradation rates in grab activated sludge inocula and the biodegradation rates used in previously developed QSBR models for four test chemicals at two different inocula concentration. Diluted refers to 10 fold dilution, whereas undiluted refers to raw activated sludge

The QSBR model calibration approach used in this study is valid for a specific studied condition, however the environmental systems are dynamic, and vary both in space and time. There is a need for a robust calibration method that enables the use of QSBR model to reliably predict the degradation of chemicals in any environmental compartment. It should be noted that, regardless of different starting chemical concentration of studied chemicals at a given inocula concentration (**Figure 7-17**), the measured biodegradation rates for the chemicals in the QSBR model might have a defined pattern of space, as indicated by significant Pearson correlation coefficient. Nevertheless, there is a need of more experimental rates at different starting chemical concentration in order to accurately identify the space of degradation rates in the QSBR models for chemicals. This might provide an avenue for robust calibration of QSBR model and ultimately can be used to predict the biodegradation rate of chemicals in dynamic environment.

Chapter 8

Conclusions, broader implication and recommendations for future work

Chapter 8. Conclusions, broader implication and recommendations for future work

8.1 Conclusions

This work provides the basis for a rational framework to establish and calibrate QSBR models that can be applied to predict biodegradation rates and ultimately the half-life of a given chemical, which is the endpoint used for persistence assessment of chemicals. In addition, this thesis also provides an insight on the microbial ecology of putative degraders for a given environment and their relationship with chemical biodegradation.

Overall conclusions:

- Theoretical biodegradation rates of chemicals (especially mono-aromatic chemicals) can be predicted with good accuracy based on chemical structures that provide a mechanistic understanding of biodegradation, using the QSBR model developed in accordance with OECD principles
- The above mentioned QSBR model was quantitatively proved using the biodegradation rates derived from pure and mixed culture experiments that demonstrated that the same rank prioritisation of biodegradation rates existed even if the absolute rates were different.
- Experimentally determined biodegradation rates can be used to calibrate such QSBRs
- A rigorous carbon mass balance model and stoichiometric equations for chemical biodegradation can ensure the reliability of measured biodegradation rates and address the uncertainty associated with the experimental determination of biodegradation rates.
- The primer pair targeting catabolic enzymes involved in rate limiting steps of the biodegradation pathway of aromatic chemical can be used as a biomarker to screen and quantify the putative chemical degraders and their activities in the inocula.
- The putative degraders can gain a competitive advantage when other taxa are disadvantaged in the presence of given specific chemical contaminant.
- Microbial degradation of chemical in natural or engineered eco-system will bring about the enrichment of putative chemical degraders, however might not necessarily influence the overall biomass count of the system
- There is a linear relationship between biodegradation rate of chemicals and ratio of initial putative degrader abundance and starting chemical concentration in the environment.
- The biodegradation of a chemical in an environment must be viewed with respect to both intrinsic properties of chemicals (e.g., chemical concentration) and inocula

characteristics (i.e. probability of encountering putative degraders of the test chemical within the test inocula).

8.2 Broader implication of the current research

QSBR models provide an opportunity to predict biodegradation rates of chemicals, which ultimately form the basis of prioritizing chemicals that pose the greatest risk to the environment and humans. However, the main concern with respect to the utilization of such models is the quality of existing biodegradation data. It has been widely accepted that a model is only good, if input data unambiguously link the model and experimental work. Furthermore, it has to be noted that majority of existing QSBR, including mine, are built with the data [e.g., Ready Biodegradability Tests (RBTs) data] from different biodegradation databases, which themselves are another QSBR. The existing regulatory screening test (i.e. RBTs) exhibits a number of limitations that are particularly pertinent to their use in persistence assessment (Kowalczyk *et al.*, 2015b). Therefore, the use of RBTs results fundamentally undermines current QSBR efforts. Models developed using such biodegradation data require further verification against experimentally determined biodegradation rates. I believe that, a traditional biodegradation test for few representative chemicals among the set of chemicals used in QSBR model development can be performed with inocula simulating the environmental compartment, and the estimated experimental biodegradation rate can be ultimately used to verify such QSBR models. A simple approach used in this study could form the basis for defining rate boundaries for calibration against the QSBR data
Regulatory bodies view degradation and persistence as properties intrinsic to the chemical in question. However, results of biodegradation assays for different test chemicals in activated sludge grab samples demonstrated that chemical structure, their concentration and inocula concentration (and characteristics) have significant influence on chemicals biodegradation rate. Results from this study suggest that, the existing approach used in current regulatory biodegradation testing methods needs modification, and should consider both intrinsic (chemical properties) and extrinsic properties (especially inocula characteristics) when determining chemical persistence. In this study, a traditional biodegradation test method was combined with 16S rRNA amplicon sequencing, functional genes, transcriptomic and cell density analysis of the inocula. The combined results from these studies, not only provided understanding on the relationship between microbial diversity and biodegradation outcomes, but also helped in evaluating the occurrence and abundance of putative chemical degraders and the bacterial variation between inocula. Indeed, these results will potentially facilitate in design of robust methods for biodegradation studies and enable better prediction of biodegradation. The latter was further verified by one of the results in this study, where

chemical fate was significantly correlated with the ratio of specific degrader abundance and chemical initial concentration. Furthermore, the gene-targeted metagenomics (i.e. identification and quantification of key catabolic genes with degenerate primers) used in this study can be used as a tool to study the diversity of key catabolic genes (e.g., oxygenases) involved in the turnover of aromatic chemicals in the environment. Such genes are biomarker genes that can be used to assess the biodegradation potential of an inoculum, and therefore prediction of the chemical fate in the environment.

Introduction of new chemicals into the environment brings about several changes within the microbial community, such as adaptation and ecological changes. Such an event normally leads to the development of novel chemical biodegradation pathways and might potentially favour the proliferation of key chemical degrading bacteria. The method used in this study combined with proteomics and metabolomics analyses form the basis for evaluating the microbial community functions, detection and identification of existing and novel enzymes and metabolites involved in the biodegradation of chemical of interest.

Therefore, in broad context, the molecular ecology methods (i.e. omics') used in this study form the basis for assessing the chemical's persistence, since these methods provide deep insight into the ongoing biochemical processes in the environment and their consequences for ecosystems.

8.3 Future outlook

A number of research questions are yet to be answered going forward from this study and should form a basis for future research on using 'omics' tools towards assessing and predicting the persistence of chemicals in both natural and engineered ecosystems.

The results from this study verified the applicability of using key catabolic genes as molecular markers in the assessment of inocula (activated sludge) biodegradation potentials and hence in the prediction of test chemical's fate. Similar studies need to be performed with inocula from other environmental systems (e.g., surface water, soil and sediments, marine water) in order to have an insight on the applicability of this approach in a broader context. Indeed, this will allow reliable classification of those chemicals that can degrade rapidly in the environment and thus prediction of chemical persistence in the environment.

The concentration of a chemical is one of the principle factors that affect chemical behaviour in the environment, and ultimately, its susceptibility to microbial attack and degradation kinetics. This study only assessed rates of biodegradation and impacts on microbial community structures at chemical concentrations generally far in excess of those likely to present in natural and engineered ecosystems. Consideration should be given to study the influence of such environmentally relevant concentrations on microbial community responses

and also on the chemical fate. For instance, the concentration of specific enzymes and biodegradation pathway to be induced in the bacterial cells for chemical degradation is greatly influenced by the chemical concentration in the environment.

This study covered the fairly narrow range of chemicals (simple aromatic chemicals). To verify the reliability of using functional gene metagenomics as an approach for predicting the persistence of chemicals in the environment, studies should be performed with chemicals having broad structural diversity and complexity, provided that the biodegradation pathways for that chemical is well established.

The biodegradation data for all the studied chemicals mostly fitted first order kinetic better than other kinetic models (second and zero order). Whereas, based on the observed change in the total biomass abundance over the duration of experiments, the biodegradation kinetics should mostly follow pseudo-first order kinetics. However, this study revealed that rare taxa (i.e. putative chemical degraders) were important niche in the microbial community, got enriched, and were responsible for the degradation of chemicals in the activated sludge systems. On the other hand, there was significant correlation between apparent first-order rates and the ratio of X_0 (initial abundance of competent putative chemical degraders) and C_0 (initial spiked chemical concentration) (i.e. X_0/C_0). These results suggest that using conventional kinetics equations (e.g., first order kinetics) might not be always realistic to determine the degradation kinetics of chemical in the environment. Thus, there is a need of extensive modelling, and a more sampling regime to cover total biomass and putative chemical degraders abundance at each point, to accurately describe the kinetics of biodegradation.

Reference

Abuhamed, T., Bayraktar, E., Mehmetoğlu, T. and Mehmetoğlu, Ü. (2004) 'Kinetics model for growth of Pseudomonas putida F1 during benzene, toluene and phenol biodegradation', *Process Biochemistry*, 39(8), pp. 983-988.

Agarry, S.E. and Ogunleye, O.O. (2012) 'Factorial designs application to study enhanced bioremediation of soil artificially contaminated with weathered bonny light crude oil through biostimulation and bioaugmentation strategy', *Journal of Environmental Protection*, 3(08), p. 748.

Alexander, M. and Lustigman, B.K. (1966) 'Effect of chemical structure on microbial degradation of substituted benzenes', *Journal of Agricultural and Food Chemistry*, 14(4), pp. 410-413.

Alfreider, A., Vogt, C. and Babel, W. (2003) 'Expression of chlorocatechol 1, 2-dioxygenase and chlorocatechol 2, 3-dioxygenase genes in chlorobenzene-contaminated subsurface samples', *Applied and environmental microbiology*, 69(3), pp. 1372-1376.

Annadurai, G., Balan, S.M. and Murugesan, T. (2000) 'Design of experiments in the biodegradation of phenol using immobilized Pseudomonas pictorum (NICM-2077) on activated carbon', *Bioprocess Engineering*, 22(2), pp. 101-107.

APHA (2005) 'Standard methods for the examination of water and wastewater', *American Public Health Association (APHA): Washington, DC, USA*.

Aranda, C., Godoy, F., Becerra, J., Barra, R. and Martínez, M. (2003) 'Aerobic secondary utilization of a non-growth and inhibitory substrate 2, 4, 6-trichlorophenol by Sphingopyxis chilensis S37 and Sphingopyxis-like strain S32', *Biodegradation*, 14(4), pp. 265-274.

Arnosti, C. (2011) 'Microbial extracellular enzymes and the marine carbon cycle', *Annual review of marine science*, 3, pp. 401-425.

Arnot, J., Gouin, T. and Mackay, D. (2005) *Development and application of models of chemical fate in Canada Practical methods for estimating environmental biodegradation rates, Canadian Environmental Modeling Network Report No. 200503*. [Online]. Available at: http://www.trentu.ca/
academic/aminss/envmodel/CEMNReport200503.pdf.

Aronson, D., Boethling, R., Howard, P. and Stiteler, W. (2006) 'Estimating biodegradation half-lives for use in chemical screening', *Chemosphere*, 63(11), pp. 1953-1960.

Arora, P.K. and Bae, H. (2014) 'Bacterial degradation of chlorophenols and their derivatives', *Microbial Cell Factories*, 13(1), p. 31.

Assinder, S.J. and Williams, P.A. (1990) 'The TOL plasmids: determinants of the catabolism of toluene and the xylenes', *Advances in microbial physiology*, 31, pp. 1-69.

Bacosa, H.P. and Inoue, C. (2015) 'Polycyclic aromatic hydrocarbons (PAHs) biodegradation potential and diversity of microbial consortia enriched from tsunami sediments in Miyagi, Japan', *Journal of hazardous materials*, 283, pp. 689-697.

Baker, J.R., Gamberger, D., Mihelcic, J.R. and Sabljic, A. (2004) 'Evaluation of artificial intelligence based models for chemical biodegradability prediction', *Molecules*, 9(12), pp. 989-1003.

Baldwin, B.R., Nakatsu, C.H. and Nies, L. (2003) 'Detection and enumeration of aromatic oxygenase genes by multiplex and real-time PCR', *Applied and Environmental Microbiology*, 69(6), pp. 3350-3358.

Ballabio, D., Manganaro, A., Consonni, V., Mauri, A. and Todeschini, R. (2009) 'Introduction to MOLE DB-on-line molecular descriptors database', *MATCH Commun Math Comput Chem*, 62, pp. 199-207.

Bamforth, S.M. and Singleton, I. (2005) 'Bioremediation of polycyclic aromatic hydrocarbons: current knowledge and future directions', *Journal of chemical technology and biotechnology*, 80(7), pp. 723-736.

Banerjee, S., Howard, P.H., Rosenberg, A.M., Dombrowski, A.E., Sikka, H. and Tullis, D.L. (1984) 'Development of a general kinetic model for biodegradation and its application to chlorophenols and related compounds', *Environmental science & technology*, 18(6), pp. 416-422.

Barraza-Villarreal, A., Escamilla-Nunez, M.C., Hernández-Cadena, L., Texcalac-Sangrador, J.L., Sienra-Monge, J.J., Del Río-Navarro, B.E., Cortez-Lugo, M., Sly, P.D. and Romieu, I. (2011) 'Elemental carbon exposure and lung function in schoolchildren from Mexico City', *European Respiratory Journal*, 38(3), pp. 548-552.

Battersby, N.S. (1990) 'A review of biodegradation kinetics in the aquatic environment', *Chemosphere*, 21(10-11), pp. 1243-1284.

Bellanger, M., Demeneix, B., Grandjean, P., Zoeller, R.T. and Trasande, L. (2015) 'Neurobehavioral deficits, diseases, and associated costs of exposure to endocrine-disrupting chemicals in the European Union', *The Journal of Clinical Endocrinology & Metabolism*, 100(4), pp. 1256-1266.

Benotti, M.J., Trenholm, R.A., Vanderford, B.J., Holady, J.C., Stanford, B.D. and Snyder, S.A. (2008) 'Pharmaceuticals and endocrine disrupting compounds in US drinking water', *Environmental science & technology*, 43(3), pp. 597-603.

Berg, U.T. and Nyholm, N. (1996) 'Biodegradability simulation studies in semicontinuous activated sludge reactors with low (μg/L range) and standard (ppm range) chemical concentrations', *Chemosphere*, 33(4), pp. 711-735.

Berney, M., Hammes, F., Bosshard, F., Weilenmann, H.-U. and Egli, T. (2007) 'Assessment and interpretation of bacterial viability by using the LIVE/DEAD BacLight Kit in combination with flow cytometry', *Applied and environmental microbiology*, 73(10), pp. 3283-3290.

Bhandari, R., Xiao, J. and Shankar, A. (2013) 'Urinary bisphenol A and obesity in US children', *American journal of epidemiology*, 177(11), pp. 1263-1270.

Blais, J.M. (2005) 'Biogeochemistry of persistent bioaccumulative toxicants: processes affecting the transport of contaminants to remote areas', *Canadian Journal of Fisheries and Aquatic Sciences*, 62(1), pp. 236-243.

Brankatschk, R., Bodenhausen, N., Zeyer, J. and Bürgmann, H. (2012) 'Simple absolute quantification method correcting for quantitative PCR efficiency variations for microbial community samples', *Applied and environmental microbiology*, 78(12), pp. 4481-4489.

Brown, M.R., Hands, C.L., Coello-Garcia, T., Sani, B.S., Ott, A.I.G., Smith, S.J. and Davenport, R.J. (2018) *Flow cytometric quantification of total bacterial numbers in activated sludge*.

Bu, Q., Shi, X., Yu, G., Huang, J. and Wang, B. (2016) 'Assessing the persistence of pharmaceuticals in the aquatic environment: Challenges and needs', *Emerging Contaminants*, 2(3), pp. 145-147.

Bugg, T., Foght, J.M., Pickard, M.A. and Gray, M.R. (2000) 'Uptake and active efflux of polycyclic aromatic hydrocarbons by Pseudomonas fluorescens LP6a', *Applied and environmental microbiology*, 66(12), pp. 5387-5392.

Bustin, S.A., Benes, V., Garson, J.A., Hellemans, J., Huggett, J., Kubista, M., Mueller, R., Nolan, T., Pfaffl, M.W. and Shipley, G.L. (2009) 'The MIQE guidelines: minimum information for publication of quantitative real-time PCR experiments', *Clinical chemistry*, 55(4), pp. 611-622.

Caporaso, J.G., Kuczynski, J., Stombaugh, J., Bittinger, K., Bushman, F.D., Costello, E.K., Fierer, N., Peña, A.G., Goodrich, J.K. and Gordon, J.I. (2010) 'QIIME allows analysis of high-throughput community sequencing data', *Nature methods*, 7(5), pp. 335-336.

Carpenter, D.O. and Bushkin-Bedient, S. (2013) 'Exposure to chemicals and radiation during childhood and risk for cancer later in life', *Journal of Adolescent Health*, 52(5), pp. S21-S29.

Cavalca, L., Hartmann, A., Rouard, N. and Guy, S. (1999) 'Diversity of tfdC genes: distribution and polymorphism among 2, 4-dichlorophenoxyacetic acid degrading soil bacteria', *FEMS microbiology ecology*, 29(1), pp. 45-58.

Cébron, A., Norini, M.-P., Beguiristain, T. and Leyval, C. (2008) 'Real-time PCR quantification of PAH-ring hydroxylating dioxygenase (PAH-RHD α) genes from Gram positive and Gram negative bacteria in soil and sediment samples', *Journal of Microbiological Methods*, 73(2), pp. 148-159.

Ceriani, L., Papa, E., Kovarich, S., Boethling, R. and Gramatica, P. (2015) 'Modeling ready biodegradability of fragrance materials', *Environmental toxicology and chemistry*, 34(6), pp. 1224-1231.

Cerniglia, C.E. (1993) 'Biodegradation of polycyclic aromatic hydrocarbons', *Current opinion in biotechnology*, 4(3), pp. 331-338.

Chaudhry, G.R. and Chapalamadugu, S. (1991) 'Biodegradation of halogenated organic compounds', *Microbiological reviews*, 55(1), pp. 59-79.

Chen, J., Li, S., Xu, B., Su, C., Jiang, Q., Zhou, C., Jin, Q., Zhao, Y. and Xiao, M. (2017) 'Characterization of Burkholderia sp. XTB-5 for Phenol Degradation and Plant Growth Promotion and Its Application in Bioremediation of Contaminated Soil', *Land Degradation & Development*, 28(3), pp. 1091-1099.

Chen, Y., Lin, C.-J., Jones, G., Fu, S. and Zhan, H. (2009) 'Enhancing biodegradation of wastewater by microbial consortia with fractional factorial design', *Journal of hazardous materials*, 171(1), pp. 948-953.

Chirico, N. and Gramatica, P. (2011) 'Real external predictivity of QSAR models: how to evaluate it? Comparison of different validation criteria and proposal of using the concordance correlation coefficient', *Journal of chemical information and modeling*, 51(9), pp. 2320-2335.

Choi, E.J., Jin, H.M., Lee, S.H., Math, R.K., Madsen, E.L. and Jeon, C.O. (2013) 'Comparative genomic analysis and benzene, toluene, ethylbenzene, and o-, m-, and p-xylene (BTEX) degradation pathways of Pseudoxanthomonas spadix BD-a59', *Applied and environmental microbiology*, 79(2), pp. 663-671.

Clarke, K.R. and Warwick, R.M. (2001) 'A further biodiversity index applicable to species lists: variation in taxonomic distinctness', *Marine ecology Progress series*, 216, pp. 265-278.

Clesceri, L.S.G., Eaton, A.E., Rice, A.D., Franson, E.W. and Mary Ann, H. (2005) *Standard methodsfor the examination of water and wastewater* (0875530478).

Codru, N., Schymura, M.J., Negoita, S., Rej, R., Carpenter, D.O. and Akwesasne Task Force on the, E. (2007) 'Diabetes in relation to serum levels of polychlorinated biphenyls and chlorinated pesticides in adult Native Americans', *Environmental health perspectives*, 115(10), p. 1442.

Colborn, T., vom Saal, F.S. and Soto, A.M. (1993) 'Developmental effects of endocrine-disrupting chemicals in wildlife and humans', *Environmental health perspectives*, 101(5), p. 378.

Copley, S.D. (2009) 'Evolution of efficient pathways for degradation of anthropogenic chemicals', *Nature chemical biology*, 5(8), pp. 559-566.

Cupul-Uicab, L.A., Skjaerven, R., Haug, K., Melve, K.K., Engel, S.M. and Longnecker, M.P. (2012) 'In utero exposure to maternal tobacco smoke and subsequent obesity, hypertension, and gestational diabetes among women in the MoBa cohort', *Environmental health perspectives*, 120(3), p. 355.

Daughton, C.G. and Ternes, T.A. (1999) 'Pharmaceuticals and personal care products in the environment: agents of subtle change?', *Environmental health perspectives*, 107(Suppl 6), p. 907.

de Menezes, A., Clipson, N. and Doyle, E. (2012) 'Comparative metatranscriptomics reveals widespread community responses during phenanthrene degradation in soil', *Environmental microbiology*, 14(9), pp. 2577-2588.

DeSantis, T.Z., Hugenholtz, P., Larsen, N., Rojas, M., Brodie, E.L., Keller, K., Huber, T., Dalevi, D., Hu, P. and Andersen, G.L. (2006) 'Greengenes, a chimera-checked 16S rRNA gene database and workbench compatible with ARB', *Applied and environmental microbiology*, 72(7), pp. 5069-5072.

Devers, M., Soulas, G. and Martin-Laurent, F. (2004) 'Real-time reverse transcription PCR analysis of expression of atrazine catabolism genes in two bacterial strains isolated from soil', *Journal of Microbiological Methods*, 56(1), pp. 3-15.

Devinyak, O., Havrylyuk, D. and Lesyk, R. (2014) '3D-MoRSE descriptors explained', *Journal of Molecular Graphics and Modelling*, 54, pp. 194-203.

Diamanti-Kandarakis, E., Bourguignon, J.-P., Giudice, L.C., Hauser, R., Prins, G.S., Soto, A.M., Zoeller, R.T. and Gore, A.C. (2009) 'Endocrine-disrupting chemicals: an Endocrine Society scientific statement', *Endocrine reviews*, 30(4), pp. 293-342.

DOE (2002) *DOE FAQ Alert Electronic Newsletter*. Available at: https://www.statease.com/news/faqalert2-4.html (Accessed: 10/11/2017).

Dolfing, J. and Janssen, D.B. (1994) 'Estimates of Gibbs free energies of formation of chlorinated aliphatic compounds', *Biodegradation*, 5(1), pp. 21-28.

Donohue, K.M., Miller, R.L., Perzanowski, M.S., Just, A.C., Hoepner, L.A., Arunajadai, S., Canfield, S., Resnick, D., Calafat, A.M. and Perera, F.P. (2013) 'Prenatal and postnatal bisphenol A exposure and asthma development among inner-city children', *Journal of Allergy and Clinical Immunology*, 131(3), pp. 736-742. e6.

Dorak, M.T. (2006) 'Real-time PCR (Advanced methods series)', *Leong SL, Hocking AD & Scott ES (2006) Survival and growth of*.

Dragon *List of molecular descriptors calculated by Dragon*. Available at: http://www.talete.mi.it/products/dragon_molecular_descriptor_list.pdf (Accessed: 22/05/2015).

Dragon (*(Software for Molecular Descriptor Caluclation), Version 6.0-2014, Talete srl. Milano, Italy, 2014, <http://www.talete.mi.it >(assessed March 2015)*. [Online]. Available at: http://www.talete.mi.it (assessed March 2015).

EC (2006) *EU REACH Regulation (EC) No 1907/2006*. Available at: http://www.chemsafetypro.com/Topics/EU/REACH_Regulation_EC_No_1907_2006.html.

EC (2008) *Priority substances under the Water Framework Directive*. Available at: http://ec.europa.eu/environment/water/water-dangersub/pri_substances.htm#dir_prior (Accessed: 01/01/2018).

EC (2011) *Common Implementation Strategy for the Water Framework Directive (2000/60/EC)*. [Online]. Available at: https://circabc.europa.eu/sd/a/0cc3581b-5f65-4b6f-91c6-433a1e947838/TGDEQS%20CIS-WFD%2027%20EC%202011.pdf.

ECHA (2008) *Guidance on information requirements and chemical safety assessment. Chapter R.11: PBT assessment' Agency, E.C. European Chemicals Agency*.

ECHA (2012) *Guidance on information requirements and chemical safety assessment. Chapter R.7b: Endpoint specific guidance.* . [Online]. Available at: http://echa.europa.eu/.

Practical guide; How to use and report (Q)SARs.

ECHA (2017) *Guidance on Information Requirements and Chemical Safety Assessment*. Available at: https://echa.europa.eu/information-on-chemicals/registered-substances (Accessed: 23/06/2017).

eChemPortal. Available at: http://www.echemportal.org/echemportal/index?pageID=0&request_locale=en (Accessed: March 2015).

Edgar, R.C. (2010) 'Search and clustering orders of magnitude faster than BLAST', *Bioinformatics*, 26(19), pp. 2460-2461.

Eggen, R.I.L., Hollender, J., Joss, A., Schärer, M. and Stamm, C. (2014) 'Reducing the discharge of micropollutants in the aquatic environment: the benefits of upgrading wastewater treatment plants'. ACS Publications.

El-Naas, M.H., Mousa, H.A. and El Gamal, M. (2017) 'Microbial Degradation of Chlorophenols', in *Microbe-Induced Degradation of Pesticides*. Springer, pp. 23-58.

Elazhari-Ali, A., Singh, A.K., Davenport, R.J., Head, I.M. and Werner, D. (2013) 'Biofuel components change the ecology of bacterial volatile petroleum hydrocarbon degradation in aerobic sandy soil', *Environmental pollution*, 173, pp. 125-132.

Elersek, T., Milavec, S., Korošec, M., Brezovsek, P., Negreira, N., Zonja, B., de Alda, M.L., Barceló, D., Heath, E. and Ščančar, J. (2016) 'Toxicity of the mixture of selected antineoplastic drugs against aquatic primary producers', *Environmental Science and Pollution Research*, 23(15), pp. 14780-14790.

EPA (2003) ' EPA Appendix A to 40 CFR, Part 423–126 Priority Pollutants'. Available at: http://www.epa.gov/region01/npdes/permits/generic/prioritypollutants.pdf.

EPA, U. (2011) 'EPI Suite TM, Version 4.0. '.

Falkenmark, M. and Rockström, J. (2004) *Balancing water for humans and nature: the new approach in ecohydrology*. Earthscan.

Fannin, T.E., Marcus, M.D., Anderson, D.A. and Bergman, H.L. (1981) 'Use of a fractional factorial design to evaluate interactions of environmental factors affecting biodegradation rates', *Applied and environmental microbiology*, 42(6), pp. 936-943.

Fent, K., Weston, A.A. and Caminada, D. (2006) 'Ecotoxicology of human pharmaceuticals', *Aquatic toxicology*, 76(2), pp. 122-159.

Finley, S.D., Broadbelt, L.J. and Hatzimanikatis, V. (2009) 'Thermodynamic analysis of biodegradation pathways', *Biotechnology and bioengineering*, 103(3), pp. 532-541.

Foladori, P., Bruni, L., Tamburini, S. and Ziglio, G. (2010) 'Direct quantification of bacterial biomass in influent, effluent and activated sludge of wastewater treatment plants by using flow cytometry', *Water Res*, 44(13), pp. 3807-18.

Futamata, H., Harayama, S. and Watanabe, K. (2001) 'Group-specific monitoring of phenol hydroxylase genes for a functional assessment of phenol-stimulated trichloroethylene bioremediation', *Applied and environmental microbiology*, 67(10), pp. 4671-4677.

Gajski, G., Gerić, M., Žegura, B., Novak, M., Nunić, J., Bajrektarević, D., Garaj-Vrhovac, V. and Filipič, M. (2016) 'Genotoxic potential of selected cytostatic drugs in human and zebrafish cells', *Environmental Science and Pollution Research*, 23(15), pp. 14739-14750.

Gamberger, D., Horvatić, D., Sekušak, S. and Sabljić, A. (1996a) 'Applications of experts' judgement to derive structure-biodegradation relationships', *Environmental Science and Pollution Research*, 3(4), pp. 224-228.

Gamberger, D., Sekušak, S., Medven, Ž. and Sabljić, A. (1996b) 'Application of Artificial Intelligence in Biodegradation Modelling', in Peijnenburg, W.G.M. and Damborský, J. (eds.) *Biodegradability Prediction*. Springer Netherlands, pp. 41-50.

Gao, J., Ellis, L. and Wackett, L.P. (2010) 'The University of Minnesota biocatalysis/biodegradation database: improving public access', *Nucleic acids research*, 38(suppl_1), pp. D488-D491.

Ghosal, D., Ghosh, S., Dutta, T.K. and Ahn, Y. (2016) 'Current state of knowledge in microbial degradation of polycyclic aromatic hydrocarbons (PAHs): a review', *Frontiers in microbiology*, 7.

Goldberg, E.D., Bourne, W.R.P., Boucher, E.A. and Preston, A. (1975) 'Synthetic organohalides in the sea', *Proceedings of the Royal Society of London B: Biological Sciences*, 189(1096), pp. 277-289.

Goodhead, A.K., Head, I.M., Snape, J.R. and Davenport, R.J. (2014) 'Standard inocula preparations reduce the bacterial diversity and reliability of regulatory biodegradation tests', *Environmental Science and Pollution Research*, 21(16), pp. 9511-9521.

Goossens, H., Ferech, M., Vander Stichele, R., Elseviers, M. and Group, E.P. (2005) 'Outpatient antibiotic use in Europe and association with resistance: a cross-national database study', *The Lancet*, 365(9459), pp. 579-587.

Gouin, T., Cousins, I. and Mackay, D. (2004) 'Comparison of two methods for obtaining degradation half-lives', *Chemosphere*, 56(6), pp. 531-535.

Grady, C.P. (1985) 'Biodegradation: its measurement and microbiological basis', *Biotechnology and bioengineering*, 27(5), pp. 660-674.

Graham, D.W., Knapp, C.W., Christensen, B.T., McCluskey, S. and Dolfing, J. (2016) 'Appearance of β-lactam resistance genes in agricultural soils and clinical isolates over the 20th century', *Scientific reports*, 6.

Gramatica, P., Chirico, N., Papa, E., Cassani, S. and Kovarich, S. (2013) 'QSARINS: A new software for the development, analysis, and validation of QSAR MLR models', *Journal of Computational Chemistry*, 34(24), pp. 2121-2132.

Grassi, M., Kaykioglu, G., Belgiorno, V. and Lofrano, G. (2012) 'Removal of emerging contaminants from water and wastewater by adsorption process', in *Emerging compounds removal from wastewater*. Springer, pp. 15-37.

Gusmão, V.R., Chinalia, F.A., Sakamoto, I.K. and Varesche, M.B.A. (2007) 'Performance of a reactor containing denitrifying immobilized biomass in removing ethanol and aromatic hydrocarbons (BTEX) in a short operating period', *Journal of hazardous materials*, 139(2), pp. 301-309.

Haiyan, R., Shulan, J., ud din Ahmad, N., Dao, W. and Chengwu, C. (2007) 'Degradation characteristics and metabolic pathway of 17α-ethynylestradiol by Sphingobacterium sp. JCR5', *Chemosphere*, 66(2), pp. 340-346.

Hall, B.G. (2011) *Phylogenetic trees made easy: A how to manual*. Sinauer.

Halling-Sørensen, B., Nielsen, S.N., Lanzky, P.F., Ingerslev, F., Lützhøft, H.C.H. and Jørgensen, S.E. (1998) 'Occurrence, fate and effects of pharmaceutical substances in the environment-A review', *Chemosphere*, 36(2), pp. 357-393.

Hammes, F., Berney, M., Wang, Y., Vital, M., Koster, O. and Egli, T. (2008) 'Flow-cytometric total bacterial cell counts as a descriptive microbiological parameter for drinking water treatment processes', *Water Res*, 42(1-2), pp. 269-77.

Hao, O.J., Kim, M.H., Seagren, E.A. and Kim, H. (2002) 'Kinetics of phenol and chlorophenol utilization by Acinetobacter species', *Chemosphere*, 46(6), pp. 797-807.

Hassen, A., Mahrouk, M., Ouzari, H., Cherif, M., Boudabous, A. and Damelincourt, J.J. (2000) 'UV disinfection of treated wastewater in a large-scale pilot plant and inactivation of selected bacteria in a laboratory UV device', *Bioresource Technology*, 74(2), pp. 141-150.

Hauser, R., Skakkebaek, N.E., Hass, U., Toppari, J., Juul, A., Andersson, A.M., Kortenkamp, A., Heindel, J.J. and Trasande, L. (2015) 'Male reproductive disorders, diseases, and costs of exposure to endocrine-disrupting chemicals in the European Union', *The Journal of Clinical Endocrinology & Metabolism*, 100(4), pp. 1267-1277.

Head, I.M., Saunders, J.R. and Pickup, R.W. (1998) 'Microbial evolution, diversity, and ecology: a decade of ribosomal RNA analysis of uncultivated microorganisms', *Microbial ecology*, 35(1), pp. 1-21.

Healey, F. (1980a) 'Slope of the Monod equation as an indicator of advantage in nutrient competition', *Microbial Ecology*, 5(4), pp. 281-286.

Healey, F.P. (1980b) 'Slope of the Monod equation as an indicator of advantage in nutrient competition', *Microbial Ecology*, 5(4), pp. 281-286.

Heath, E., Filipič, M., Kosjek, T. and Isidori, M. (2016) 'Fate and effects of the residues of anticancer drugs in the environment', *Environmental Science and Pollution Research*, 23(15), pp. 14687-14691.

Heberer, T. (2002) 'Occurrence, fate, and removal of pharmaceutical residues in the aquatic environment: a review of recent research data', *Toxicology letters*, 131(1), pp. 5-17.

Helbling, D.E., Johnson, D.R., Honti, M. and Fenner, K. (2012) 'Micropollutant biotransformation kinetics associate with WWTP process parameters and microbial community characteristics', *Environmental science & technology*, 46(19), pp. 10579-10588.

Helguera, A.M., Combes, R.D., Gonzalez, M.P. and Cordeiro, M. (2008) 'Applications of 2D descriptors in drug design: a DRAGON tale', *Current topics in medicinal chemistry*, 8(18), pp. 1628-1655.

Hickman, G.T. and Novak, J.T. (1989) 'Relationship between subsurface biodegradation rates and microbial density', *Environmental science & technology*, 23(5), pp. 525-532.

Hollams, E.M., De Klerk, N.H., Holt, P.G. and Sly, P.D. (2014) 'Persistent effects of maternal smoking during pregnancy on lung function and asthma in adolescents', *American journal of respiratory and critical care medicine*, 189(4), pp. 401-407.

Hong, J. (1989) 'Yield coefficients for cell mass and product formation', *Biotechnology and bioengineering*, 33(4), pp. 506-507.

Hu, W.-S. (2017) *Engineering Principles in Biotechnology*. John Wiley & Sons.

Hutchinson, D.H. and Robinson, C.W. (1988) 'Kinetics of the simultaneous batch degradation of p-cresol and phenol by Pseudomonas putida', *Applied microbiology and biotechnology*, 29(6), pp. 599-604.

iitd.vlab.co.in (2012) *Estimation of growth kinetic parameters in batch fermentation* [Computer program]. Available at: iitd.vlab.co.in/?sub=63&brch=177&sim=1348&cnt=1

Itrich, N.R., McDonough, K.M., van Ginkel, C.G., Bisinger, E.C., LePage, J.N., Schaefer, E.C., Menzies, J.Z., Casteel, K.D. and Federle, T.W. (2015) 'Widespread microbial adaptation to L-glutamate-N, N-diacetate (L-GLDA) following its market introduction in a consumer cleaning product', *Environmental science & technology*, 49(22), pp. 13314-13321.

Iwai, S., Johnson, T.A., Chai, B., Hashsham, S.A. and Tiedje, J.M. (2011) 'Comparison of the specificities and efficacies of primers for aromatic dioxygenase gene analysis of environmental samples', *Applied and environmental microbiology*, 77(11), pp. 3551-3557.

Jones, O.A.H., Voulvoulis, N. and Lester, J.N. (2004) 'Potential ecological and human health risks associated with the presence of pharmaceutically active compounds in the aquatic environment', *Critical reviews in toxicology*, 34(4), pp. 335-350.

Jung, J., Philippot, L. and Park, W. (2016) 'Metagenomic and functional analyses of the consequences of reduction of bacterial diversity on soil functions and bioremediation in diesel-contaminated microcosms', *Scientific reports*, 6, p. 23012.

Kanaly, R.A., Harayama, S. and Watanabe, K. (2002) 'Rhodanobacter sp. strain BPC1 in a benzo [a] pyrene-mineralizing bacterial consortium', *Applied and environmental microbiology*, 68(12), pp. 5826-5833.

Karelson, M., Lobanov, V.S. and Katritzky, A.R. (1996) 'Quantum-chemical descriptors in QSAR/QSPR studies', *Chemical reviews*, 96(3), pp. 1027-1044.

Kennedy, K., Hall, M.W., Lynch, M.D.J., Moreno-Hagelsieb, G. and Neufeld, J.D. (2014) 'Evaluating bias of Illumina-based bacterial 16S rRNA gene profiles', *Applied and environmental microbiology*, 80(18), pp. 5717-5722.

Klopman, G., Dimayuga, M. and Talafous, J. (1994) 'META. 1. A program for the evaluation of metabolic transformation of chemicals', *Journal of Chemical Information and Computer Sciences*, 34(6), pp. 1320-1325.

Koch, R. (1982) 'Molecular connectivity and acute toxicity of environmental pollutants', *Chemosphere*, 11(9), pp. 925-931.

Kolpin, D.W., Furlong, E.T., Meyer, M.T., Thurman, E.M., Zaugg, S.D., Barber, L.B. and Buxton, H.T. (2002) 'Pharmaceuticals, hormones, and other organic wastewater contaminants in US streams, 1999− 2000: A national reconnaissance', *Environmental science & technology*, 36(6), pp. 1202-1211.

Kovács, R., Csenki, Z., Bakos, K., Urbányi, B., Horváth, Á., Garaj-Vrhovac, V., Gajski, G., Gerić, M., Negreira, N. and de Alda, M.L. (2015) 'Assessment of toxicity and genotoxicity of

low doses of 5-fluorouracil in zebrafish (Danio rerio) two-generation study', *Water research*, 77, pp. 201-212.

Kovárová-Kovar, K. and Egli, T. (1998) 'Growth kinetics of suspended microbial cells: from single-substrate-controlled growth to mixed-substrate kinetics', *Microbiology and molecular biology reviews*, 62(3), pp. 646-666.

Kowalczyk, A., Eyice, Ö., Schäfer, H., Price, O.R., Finnegan, C.J., van Egmond, R.A., Shaw, L.J., Barrett, G. and Bending, G.D. (2015a) 'Characterization of para-Nitrophenol-Degrading Bacterial Communities in River Water by Using Functional Markers and Stable Isotope Probing', *Applied and environmental microbiology*, 81(19), pp. 6890-6900.

Kowalczyk, A., Martin, T.J., Price, O.R., Snape, J.R., van Egmond, R.A., Finnegan, C.J., Schäfer, H., Davenport, R.J. and Bending, G.D. (2015b) 'Refinement of biodegradation tests methodologies and the proposed utility of new microbial ecology techniques', *Ecotoxicology and environmental safety*, 111, pp. 9-22.

Kowalczyk, A., Price, O.R., van der Gast, C.J., Finnegan, C.J., van Egmond, R.A., Schäfer, H. and Bending, G.D. (2016) 'Spatial and temporal variability in the potential of river water biofilms to degrade p-nitrophenol', *Chemosphere*, 164, pp. 355-362.

Kümmerer, K. (2010) 'Pharmaceuticals in the environment', *Annual review of environment and resources*, 35, pp. 57-75.

Kundi, M., Parrella, A., Lavorgna, M., Criscuolo, E., Russo, C. and Isidori, M. (2016) 'Prediction and assessment of ecogenotoxicity of antineoplastic drugs in binary mixtures', *Environmental Science and Pollution Research*, 23(15), pp. 14771-14779.

La Merrill, M., Cirillo, P.M., Terry, M.B., Krigbaum, N.Y., Flom, J.D. and Cohn, B.A. (2013) 'Prenatal exposure to the pesticide DDT and hypertension diagnosed in women before age 50: a longitudinal birth cohort study', *Environmental health perspectives*, 121(5), p. 594.

Larkin, M.A., Blackshields, G., Brown, N.P., Chenna, R., McGettigan, P.A., McWilliam, H., Valentin, F., Wallace, I.M., Wilm, A. and Lopez, R. (2007) 'Clustal W and Clustal X version 2.0', *bioinformatics*, 23(21), pp. 2947-2948.

Leatherland, J.F. (1992) 'Endocrine and reproductive function in Great Lakes salmon', *Chemicall Induced Alterations in Sexual and Functional Development: The Wild Life/Human Connection.*, pp. 129-145.

Lee, D.-J., Ho, K.-L. and Chen, Y.-Y. (2011) 'Degradation of cresols by phenol-acclimated aerobic granules', *Applied microbiology and biotechnology*, 89(1), pp. 209-215.

Lee, D.S., Ryu, S.H., Hwang, H.W., Kim, Y.-J., Park, M., Lee, J.R., Lee, S.-S. and Jeon, C.O. (2008) 'Pseudoxanthomonas sacheonensis sp. nov., isolated from BTEX-contaminated soil in Korea, transfer of Stenotrophomonas dokdonensis Yoon et al. 2006 to the genus Pseudoxanthomonas as Pseudoxanthomonas dokdonensis comb. nov. and emended description of the genus Pseudoxanthomonas', *International journal of systematic and evolutionary microbiology*, 58(9), pp. 2235-2240.

Lee, S.H., Jin, H.M., Lee, H.J., Kim, J.M. and Jeon, C.O. (2012) 'Complete genome sequence of the BTEX-degrading bacterium Pseudoxanthomonas spadix BD-a59', *Journal of bacteriology*, 194(2), pp. 544-544.

Lee, Y. and von Gunten, U. (2012) 'Quantitative structure-activity relationships (QSARs) for the transformation of organic micropollutants during oxidative water treatment', *Water Res*, 46(19), pp. 6177-95.

Leveau, J.H.J., KoÈnig, F., FuÈchslin, H., Werlen, C., Der Meer, V. and Roelof, J. (1999) 'Dynamics of multigene expression during catabolic adaptation of Ralstonia eutropha JMP134 (pJP4) to the herbicide 2, 4-dichlorophenoxyacetate', *Molecular microbiology*, 33(2), pp. 396-406.

Lillis, L., Clipson, N. and Doyle, E. (2010) 'Quantification of catechol dioxygenase gene expression in soil during degradation of 2, 4-dichlorophenol', *FEMS microbiology ecology*, 73(2), pp. 363-369.

Liu, Z., Xie, W., Li, D., Peng, Y., Li, Z. and Liu, S. (2016) 'Biodegradation of phenol by bacteria strain Acinetobacter calcoaceticus PA isolated from phenolic wastewater', *International journal of environmental research and public health*, 13(3), p. 300.

Loonen, H., Lindgren, F., Hansen, B., Karcher, W., Niemelä, J., Hiromatsu, K., Takatsuki, M., Peijnenburg, W., Rorije, E. and Struijś, J. (1999) 'Prediction of biodegradability from chemical structure: modeling of ready biodegradation test data', *Environmental toxicology and chemistry*, 18(8), pp. 1763-1768.

Louie, T.M., Webster, C.M. and Xun, L. (2002) 'Genetic and biochemical characterization of a 2, 4, 6-trichlorophenol degradation pathway in Ralstonia eutropha JMP134', *Journal of bacteriology*, 184(13), pp. 3492-3500.

Lvovitch, M.I. (1973) 'The global water balance', *Eos, Transactions American Geophysical Union*, 54(1), pp. 28-53.

Mangse, G. (2016) *Investigating the effects of biochar and activated carbon amendment on the microbial community response in a volatile petroleum hydrocarbon -contaminated gravelly sand*. Newcastle University.

Mansouri, K., Ringsted, T., Ballabio, D., Todeschini, R. and Consonni, V. (2013) 'Quantitative structure–activity relationship models for ready biodegradability of chemicals', *Journal of chemical information and modeling*, 53(4), pp. 867-878.

Marangoni, C., Furigo Jr, A. and Aragão, G.M. (2001) 'The influence of substrate source on the growth of Ralstonia eutropha, aiming at the production of polyhydroxyalkanoate', *Brazilian Journal of Chemical Engineering*, 18(2), pp. 175-180.

Mars, A.E., Houwing, J., Dolfing, J. and Janssen, D.B. (1996) 'Degradation of Toluene and Trichloroethylene by Burkholderia cepacia G4 in Growth-Limited Fed-Batch Culture', *Applied and environmental microbiology*, 62(3), pp. 886-891.

Martin, T.J. (2014) *THE INFLUENCE OF MICROBIAL INOCULA ON BIODEGRADATION OUTCOME TOWARDS ENHANCED REGULATORY ASSESSMENTS*. Newcastle University.

Martin, T.J., Goodhead, A.K., Acharya, K., Head, I.M., Snape, J.R. and Davenport, R.J. (2017a) 'A high-throughput biodegradation-screening test to prioritise and evaluate chemical biodegradability', *Environmental Science & Technology*.

Martin, T.J., Snape, J.R., Bartram, A., Robson, A., Acharya, K. and Davenport, R.J. (2017b) 'Environmentally relevant inocula concentrations improve the reliability of persistent assessments in biodegradation screening tests', *Environmental Science & Technology*.

Martin, T.J., Snape, J.R., Bartram, A., Robson, A., Acharya, K. and Davenport, R.J. (2017c) 'Environmentally Relevant Inoculum Concentrations Improve the Reliability of Persistent Assessments in Biodegradation Screening Tests', *Environmental Science & Technology*, 51(5), pp. 3065-3073.

Matasci, R., Weston, R., Lau, P., Cruver, J., Marek, S. and Tomowich, D. (1999) *Wastewater Technology Fact Sheet: Ultraviolet Disinfection, United States Environmental Protection Agency, Office of Water, Washington, DC*. EPA 832-F-99-064.

Mauri, A., Consonni, V., Pavan, M. and Todeschini, R. (2006) 'Dragon software: An easy approach to molecular descriptor calculations', *Match*, 56(2), pp. 237-248.

McCarty, P.L. (2012) *Environmental biotechnology: principles and applications*. Tata McGraw-Hill Education.

Mergaert, J., Cnockaert, M.C. and Swings, J. (2003) 'Thermomonas fusca sp. nov. and Thermomonas brevis sp. nov., two mesophilic species isolated from a denitrification reactor with poly (ε-caprolactone) plastic granules as fixed bed, and emended description of the genus Thermomonas', *International journal of systematic and evolutionary microbiology*, 53(6), pp. 1961-1966.

Mesarch, M.B., Nakatsu, C.H. and Nies, L. (2000) 'Development of catechol 2, 3-dioxygenase-specific primers for monitoring bioremediation by competitive quantitative PCR', *Applied and Environmental Microbiology*, 66(2), pp. 678-683.

Meynet, P., Head, I.M., Werner, D. and Davenport, R.J. (2015) 'Re-evaluation of dioxygenase gene phylogeny for the development and validation of a quantitative assay for environmental aromatic hydrocarbon degraders', *FEMS microbiology ecology*, 91(6), p. fiv049.

Midoro-Horiuti, T., Lin, Y., Nauduri, D., Kaphalia, B.S. and Goldblum, R.M. (2009) 'Prenatal Exposure of Biophenol A (BPA) Enhances Allergic Sensitization', *Journal of Allergy and Clinical Immunology*, 123(2), p. S149.

Mikolajczyk, A., Gajewicz, A., Rasulev, B., Schaeublin, N., Maurer-Gardner, E., Hussain, S., Leszczynski, J. and Puzyn, T. (2015) 'Zeta Potential for Metal Oxide Nanoparticles: A Predictive Model Developed by a Nano-Quantitative Structure–Property Relationship Approach', *Chemistry of Materials*, 27(7), pp. 2400-2407.

Min, H. (2004) 'Characterization of a strain of Sphingobacterium sp. and its degradation to herbicide mefenacet', *Journal of Environmental Sciences*, 16(2), pp. 343-347.

Minitab17 (2010) *[Statistical Computer software]* [Computer program]. State College, PA: Minitab, Inc. Available at: www.minitab.com.

Moccia, R.D., Fox, G.A. and Britton, A. (1986) 'A quantitative assessment of thyroid histopathology of herring gulls (Larus argentatus) from the Great Lakes and a hypothesis on the causal role of environmental contaminants', *Journal of Wildlife Diseases*, 22(1), pp. 60-70.

Moccia, R.D., Leatherland, J.F. and Sonstegard, R.A. (1981) 'Quantitative interlake comparison of thyroid pathology in Great Lakes coho (Oncorhynchus kisutch) and chinook (Oncorhynchus tschawytscha) salmon', *Cancer Research*, 41(6), pp. 2200-2210.

Monod, J. (1949) 'The growth of bacterial cultures', *Annual Reviews in Microbiology*, 3(1), pp. 371-394.

Moon, K., Guallar, E. and Navas-Acien, A. (2012) 'Arsenic exposure and cardiovascular disease: an updated systematic review', *Current atherosclerosis reports*, 14(6), pp. 542-555.

Moran, M.A. (2009) 'Metatranscriptomics: Eavesdropping on Complex Microbial Communities-Large-scale sequencing of mRNAs retrieved from natural communities provides insights into microbial activities and how they are regulated', *Microbe*, 4(7), p. 329.

Morley, R., Payne, C.L., Lister, G. and Lucas, A. (1995) 'Maternal smoking and blood pressure in 7.5 to 8 year old offspring', *Archives of disease in childhood*, 72(2), pp. 120-124.

Mukherjee, A.K. and Bordoloi, N.K. (2012) 'Biodegradation of benzene, toluene, and xylene (BTX) in liquid culture and in soil by Bacillus subtilis and Pseudomonas aeruginosa strains and a formulated bacterial consortium', *Environmental Science and Pollution Research*, 19(8), pp. 3380-3388.

Nendza, M., Gabbert, S., Kühne, R., Lombardo, A., Roncaglioni, A., Benfenati, E., Benigni, R., Bossa, C., Strempel, S. and Scheringer, M. (2013) 'A comparative survey of chemistry-driven in silico methods to identify hazardous substances under REACH', *Regulatory Toxicology and Pharmacology*, 66(3), pp. 301-314.

Netzeva, T.I., Worth, A.P., Aldenberg, T., Benigni, R., Cronin, M.T.D., Gramatica, P., Jaworska, J.S., Kahn, S., Klopman, G. and Marchant, C.A. (2005) 'Current status of methods for defining the applicability domain of (quantitative) structure-activity relationships', *ATLA*, 33, pp. 155-173.

Neu, H.C. (1992) 'The crisis in antibiotic resistance', *Science*, 257(5073), pp. 1064-1074.

2455/2001/EC of the European Parliament and of the Council of 20 November 2001 establishing the list of priority substances in the field of water policy and amending Directive 2000/60/EC (15).

Nolte, T.M. and Ragas, A.M.J. (2017) 'A review of quantitative structure–property relationships for the fate of ionizable organic chemicals in water matrices and identification of knowledge gaps', *Environmental Science: Processes & Impacts*, 19(3), pp. 221-246.

Norman, R.E., Ryan, A., Grant, K., Sitas, F. and Scott, J.G. (2014) 'Environmental contributions to childhood cancers', *J Environ Immunology and Toxicology*, 1(4), p. 12.

OECD guideline for testing of chemicals 302A: Inherent biodegradability: Modified SCAS test. .

OECD guideline for testing of chemicals 302C: Modified MITI test (II).

OECD (1992a) 'OECD Guideline for testing of chemicals'. Organisation for Economic Co-operation and Development, Paris.

Test No. 302B: Inherent Biodegradability: Zahn-Wellens/ EVPA Test.

OECD (1994) *OECD Guidelines for the Testing of Chemicals*. Organization for Economic.

OECD (2003) *INTRODUCTION TO THE OECD GUIDELINES FOR TESTING OF CHEMICALS SECTION 3* [Online]. Available at: http://www.oecd.org/chemicalsafety/testing/5598432.pdf (Accessed: 22/10/207).

OECD (2017). Available at: https://www.echemportal.org/echemportal/index.action (Accessed: 23/06/2017).

Okey, R.W. and Stensel, H.D. (1993) 'A QSBR development procedure for aromatic xenobiotic degradation by unacclimated bacteria', *Water environment research*, 65(6), pp. 772-780.

Okey, R.W. and Stensel, H.D. (1996) 'A QSAR-based biodegradability model—a QSBR', *Water Research*, 30(9), pp. 2206-2214.

Oki, T., Agata, Y., Kanae, S., Saruhashi, T., Yang, D. and Musiake, K. (2001) 'Global assessment of current water resources using total runoff integrating pathways', *Hydrological Sciences Journal*, 46(6), pp. 983-995.

Okpokwasili, G.C. and Nweke, C.O. (2006) 'Microbial growth and substrate utilization kinetics', *African Journal of Biotechnology*, 5(4), pp. 305-317.

Oliveira, C.S., Ordaz, A., Alba, J., Alves, M., Ferreira, E.C. and Thalasso, F. (2009) 'Determination of kinetic and stoichiometric parameters of Pseudomonas putida F1 by chemostat and in situ pulse respirometry', *Chemical Product and Process Modeling*, 4(2).

Ong, S.H., Kukkillaya, V.U., Wilm, A., Lay, C., Ho, E.X.P., Low, L., Hibberd, M.L. and Nagarajan, N. (2013) 'Species identification and profiling of complex microbial communities using shotgun Illumina sequencing of 16S rRNA amplicon sequences', *PLoS One*, 8(4), p. e60811.

Paisio, C.E., Quevedo, M.R., Talano, M.A., González, P.S. and Agostini, E. (2014) 'Application of two bacterial strains for wastewater bioremediation and assessment of phenolics biodegradation', *Environmental technology*, 35(14), pp. 1802-1810.

Palleroni, N.J., Pieper, D.H. and Moore, E.R.B. (2010) 'Microbiology of hydrocarbon-degrading Pseudomonas', in *Handbook of Hydrocarbon and Lipid Microbiology*. Springer, pp. 1787-1798.

Paris, D.F. and Wolfe, N.L. (1987) 'Relationship between properties of a series of anilines and their transformation by bacteria', *Applied and environmental microbiology*, 53(5), pp. 911-916.

Paris, D.F., Wolfe, N.L. and Steen, W.C. (1982) 'Structure-activity relationships in microbial transformation of phenols', *Applied and environmental microbiology*, 44(1), pp. 153-158.

Paris, D.F., Wolfe, N.L., Steen, W.C. and Baughman, G.L. (1983) 'Effect of phenol molecular structure on bacterial transformation rate constants in pond and river samples', *Applied and environmental microbiology*, 45(3), pp. 1153-1155.

Parks, D.H., Tyson, G.W., Hugenholtz, P. and Beiko, R.G. (2014) 'STAMP: statistical analysis of taxonomic and functional profiles', *Bioinformatics*, 30(21), pp. 3123-3124.

Parsons, J. and Govers, H. (1990a) 'Quantitative structure-activity relationships for biodegradation', *Ecotoxicology and environmental safety*, 19(2), pp. 212-227.

Parsons, J.R. and Govers, H.A.J. (1990b) 'Quantitative structure-activity relationships for biodegradation', *Ecotoxicology and environmental safety*, 19(2), pp. 212-227.

Pavan, M. and Worth, A.P. (2008) 'Review of Estimation Models for Biodegradation', *QSAR & Combinatorial Science*, 27(1), pp. 32-40.

Peijnenburg, W. (1994) 'Structure-activity relationships for biodegradation: a critical review', *Pure and applied chemistry*, 66(9), pp. 1931-1941.

Peijnenburg, W.J. and Damborský, J. (2012) *Biodegradability prediction*. Springer Science & Business Media.

Peng, A., Liu, J., Ling, W., Chen, Z. and Gao, Y. (2015) 'Diversity and distribution of 16S rRNA and phenol monooxygenase genes in the rhizosphere and endophytic bacteria isolated from PAH-contaminated sites', *Scientific reports*, 5.

Pérez-Pantoja, D., Donoso, R., Junca, H., González, B. and Pieper, D.H. (2010) 'Phylogenomics of Aerobic Bacterial Degradation of Aromatics', pp. 1355-1397.

Peters, A., Döring, A., Wichmann, H.E. and Koenig, W. (1997) 'Increased plasma viscosity during an air pollution episode: a link to mortality?', *The Lancet*, 349(9065), pp. 1582-1587.

Petropoulos, E. (2015) *INVESTIGATING THE TRUE LIMITS OF ANAEROBIC TREATMENT OF WASTEWATER AT LOW TEMPERATURE USING A COLD-ADAPTED INOCULUM*. Newcastle University.

Petrucci, R.H., Harwood, W.S. and Herring, F. (1997) 'General chemistry, principles and modern applications', *Journal of Chemical Education*, 74(5), pp. 491-491.

Pirhadi, S., Shiri, F. and Ghasemi, J.B. (2015) 'Multivariate statistical analysis methods in QSAR', *RSC Advances*, 5(127), pp. 104635-104665.

Pitter, P. and Chudoba, J. (1990) *Biodegradability of organic substance in the aquatic environment*. Boca Raton, Florida, United States: CRC Press, Inc.

Plumeier, I., Pérez-Pantoja, D., Heim, S., González, B. and Pieper, D.H. (2002) 'Importance of different tfd genes for degradation of chloroaromatics by Ralstonia eutropha JMP134', *Journal of bacteriology*, 184(15), pp. 4054-4064.

Powlowski, J. and Shingler, V. (1994) 'Genetics and biochemistry of phenol degradation byPseudomonas sp. CF600', *Biodegradation*, 5(3-4), pp. 219-236.

Puzyn, T., Leszczynska, D. and Leszczynski, J. (2009) 'Toward the development of "nano-QSARs": Advances and challenges', *Small*, 5(22), pp. 2494-2509.

Raymond, J.W., Rogers, T.N., Shonnard, D.R. and Kline, A.A. (2001) 'A review of structure-based biodegradation estimation methods', *Journal of hazardous materials*, 84(2-3), pp. 189-215.

Reardon, K.F., Mosteller, D.C. and Bull Rogers, J.D. (2000) 'Biodegradation kinetics of benzene, toluene, and phenol as single and mixed substrates for Pseudomonas putida F 1', *Biotechnology and Bioengineering*, 69(4), pp. 385-400.

Reeder, J. and Knight, R. (2010) 'Rapidly denoising pyrosequencing amplicon reads by exploiting rank-abundance distributions', *Nature methods*, 7(9), pp. 668-669.

Reijnders, P.J.H. (1986) 'Reproductive failure in common seals feeding on fish from polluted coastal waters', *Nature*, 324(6096), pp. 456-457.

Reuschenbach, P., Pagga, U. and Strotmann, U. (2003) 'A critical comparison of respirometric biodegradation tests based on OECD 301 and related test methods', *Water research*, 37(7), pp. 1571-1582.

Revathy, T., Jayasri, M.A. and Suthindhiran, K. (2015) 'Biodegradation of PAHs by Burkholderia sp. VITRSB1 isolated from marine sediments', *Scientifica*, 2015.

Rigas, F., Papadopoulou, K., Philippoussis, A., Papadopoulou, M. and Chatzipavlidis, J. (2009) 'Bioremediation of lindane contaminated soil by Pleurotus ostreatus in non sterile conditions using multilevel factorial design', *Water, air, and soil pollution*, 197(1-4), pp. 121-129.

Rolfe, M.D., Rice, C.J., Lucchini, S., Pin, C., Thompson, A., Cameron, A.D., Alston, M., Stringer, M.F., Betts, R.P. and Baranyi, J. (2012) 'Lag phase is a distinct growth phase that prepares bacteria for exponential growth and involves transient metal accumulation', *Journal of bacteriology*, 194(3), pp. 686-701.

Roy, P.P., Kovarich, S. and Gramatica, P. (2011) 'QSAR model reproducibility and applicability: A case study of rate constants of hydroxyl radical reaction models applied to polybrominated diphenyl ethers and (benzo-) triazoles', *Journal of computational chemistry*, 32(11), pp. 2386-2396.

Rücker, C. and Kümmerer, K. (2012) 'Modeling and predicting aquatic aerobic biodegradation–a review from a user's perspective', *Green Chemistry*, 14(4), pp. 875-887.

Sabljic, A. and Peijnenburg, W. (2001) 'Modeling lifetime and degradability of organic compounds in air, soil, and water systems (IUPAC Technical Report)', *Pure and Applied Chemistry*, 73(8), pp. 1331-1348.

Sánchez-Avila, J., Bonet, J., Velasco, G. and Lacorte, S. (2009) 'Determination and occurrence of phthalates, alkylphenols, bisphenol A, PBDEs, PCBs and PAHs in an industrial sewage grid discharging to a Municipal Wastewater Treatment Plant', *Science of the Total Environment*, 407(13), pp. 4157-4167.

Santisi, S., Cappello, S., Catalfamo, M., Mancini, G., Hassanshahian, M., Genovese, L., Giuliano, L. and Yakimov, M.M. (2015) 'Biodegradation of crude oil by individual bacterial strains and a mixed bacterial consortium', *Brazilian Journal of Microbiology*, 46(2), pp. 377-387.

Saterbak, A., San, K.-Y. and McIntire, L.V. (2007) *Bioengineering fundamentals*. Prentice Hall.

Scheringer, M. (2009) 'Long-range transport of organic chemicals in the environment', *Environmental Toxicology and Chemistry*, 28(4), pp. 677-690.

Schikowski, T., Mills, I.C., Anderson, H.R., Cohen, A., Hansell, A., Kauffmann, F., Krämer, U., Marcon, A., Perez, L. and Sunyer, J. (2014) 'Ambient air pollution: a cause of COPD?', *European Respiratory Journal*, 43(1), pp. 250-263.

Schirmer, M., Ijaz, U.Z., D'Amore, R., Hall, N., Sloan, W.T. and Quince, C. (2015) 'Insight into biases and sequencing errors for amplicon sequencing with the Illumina MiSeq platform', *Nucleic acids research*, 43(6), pp. e37-e37.

Schwarzenbach, R.P., Escher, B.I., Fenner, K., Hofstetter, T.B., Johnson, C.A., Von Gunten, U. and Wehrli, B. (2006) 'The challenge of micropollutants in aquatic systems', *Science*, 313(5790), pp. 1072-1077.

Sei, K., Asano, K.-I., Tateishi, N., Mori, K., Ike, M. and Fujita, M. (1999) 'Design of PCR primers and gene probes for the general detection of bacterial populations capable of degrading aromatic compounds via catechol cleavage pathways', *Journal of bioscience and bioengineering*, 88(5), pp. 542-550.

Shaw, A.K., Halpern, A.L., Beeson, K., Tran, B., Venter, J.C. and Martiny, J.B.H. (2008) 'It's all relative: ranking the diversity of aquatic bacterial communities', *Environmental microbiology*, 10(9), pp. 2200-2210.

Shaw, L.J., Beaton, Y., Glover, L.A., Killham, K. and Meharg, A.A. (1999) 'Development and characterization of a lux-modified 2, 4-dichlorophenol-degrading Burkholderia sp. RASC', *Environmental microbiology*, 1(5), pp. 393-400.

Shehata, T.E. and Marr, A.G. (1971) 'Effect of nutrient concentration on the growth of Escherichia coli', *Journal of bacteriology*, 107(1), pp. 210-216.

Shugart, G.W. (1980) 'Frequency and distribution of polygyny in Great Lakes Herring Gulls in 1978', *Condor*, pp. 426-429.

Sigma-Aldrich. Available at: http://www.sigmaaldrich.com/analytical-chromatography/vials/mininert-valves.html.

Sikarwar, A.K. and Dixit, S. (2012) 'QSAR MODELING OF PHENOL DERIVATIVES USING 3D MoRSE DESCRIPTORS AND EIGEN VALUES', *Journal of Environmental Research And Development Vol*, 6(3A).

Simkins, S. and Alexander, M. (1984) 'Models for mineralization kinetics with the variables of substrate concentration and population density', *Applied and Environmental Microbiology*, 47(6), pp. 1299-1306.

Simkins, S. and Alexander, M. (1985) 'Nonlinear estimation of the parameters of Monod kinetics that best describe mineralization of several substrate concentrations by dissimilar bacterial densities', *Applied and environmental microbiology*, 50(4), pp. 816-824.

Sly, P.D., Carpenter, D.O., Van den Berg, M., Stein, R.T., Landrigan, P.J., Brune-Drisse, M.-N. and Suk, W. (2016) 'Health consequences of environmental exposures: causal thinking in global environmental epidemiology', *Annals of global health*, 82(1), pp. 3-9.

Solomon, K.R., Baker, D.B., Richards, R.P., Dixon, K.R., Klaine, S.J., La Point, T.W., Kendall, R.J., Weisskopf, C.P., Giddings, J.M. and Giesy, J.P. (1996) 'Ecological risk assessment of atrazine in North American surface waters', *Environmental toxicology and Chemistry*, 15(1), pp. 31-76.

Spain, J.C. and Gibson, D.T. (1988) 'Oxidation of substituted phenols by Pseudomonas putida F1 and Pseudomonas sp. strain JS6', *Applied and environmental microbiology*, 54(6), pp. 1399-1404.

Stamm, C., Eggen, R.I.L., Hering, J.G., Hollender, J., Joss, A. and Schärer, M. (2015) 'Micropollutant removal from wastewater: Facts and decision-making despite uncertainty'. ACS Publications.

Steinle, P., Stucki, G., Stettler, R. and Hanselmann, K.W. (1998) 'Aerobic mineralization of 2, 6-dichlorophenol by Ralstonia sp. strain RK1', *Applied and environmental microbiology*, 64(7), pp. 2566-2571.

Stewart, J.J.P. (1994) *MOPAC 93 manual revision number 2. FUJITSU LIMITED*.

Stumm, W. and Morgan, J.J. (1970) *Aquatic chemistry; an introduction emphasizing chemical equilibria in natural waters*.

Sumpter, J.P. (2009) 'Protecting aquatic organisms from chemicals: the harsh realities', *Philosophical Transactions of the Royal Society of London A: Mathematical, Physical and Engineering Sciences*, 367(1904), pp. 3877-3894.

Sutton, N.B., Maphosa, F., Morillo, J.A., Al-Soud, W.A., Langenhoff, A.A.M., Grotenhuis, T., Rijnaarts, H.H.M. and Smidt, H. (2013) 'Impact of long-term diesel contamination on soil microbial community structure', *Applied and environmental microbiology*, 79(2), pp. 619-630.

Tamboli, D.P., Kurade, M.B., Waghmode, T.R., Joshi, S.M. and Govindwar, S.P. (2010) 'Exploring the ability of Sphingobacterium sp. ATM to degrade textile dye Direct Blue GLL, mixture of dyes and textile effluent and production of polyhydroxyhexadecanoic acid using waste biomass generated after dye degradation', *Journal of Hazardous Materials*, 182(1), pp. 169-176.

Tamura, K., Peterson, D., Peterson, N., Stecher, G., Nei, M. and Kumar, S. (2011) 'MEGA5: molecular evolutionary genetics analysis using maximum likelihood, evolutionary distance, and maximum parsimony methods', *Molecular biology and evolution*, 28(10), pp. 2731-2739.

Tan, B., Ng, C., Nshimyimana, J.P., Loh, L.L., Gin, K.Y.H. and Thompson, J.R. (2015) 'Next-generation sequencing (NGS) for assessment of microbial water quality: current progress, challenges, and future opportunities', *Frontiers in microbiology*, 6.

Thouand, G., Capdeville, B. and Block, J.C. (1996) 'Preadapted inocula for limiting the risk of errors in biodegradability tests', *Ecotoxicology and environmental safety*, 33(3), pp. 261-267.

Thouand, G., Friant, P., Bois, F., Cartier, A., Maul, A. and Block, J.C. (1995) 'Bacterial inoculum density and probability of para-nitrophenol biodegradability test response', *Ecotoxicology and environmental safety*, 30(3), pp. 274-282.

Todeschini, R. and Consonni, V. (2008) *Handbook of molecular descriptors*. John Wiley & Sons.

Trousselier, M., Bouvy, M., Courties, C. and Dupuy, C. (1997) 'Variation of carbon content among bacterial species under starvation condition', *Aquatic microbial ecology*.

Tuan, N.N., Hsieh, H.-C., Lin, Y.-W. and Huang, S.-L. (2011) 'Analysis of bacterial degradation pathways for long-chain alkylphenols involving phenol hydroxylase, alkylphenol

monooxygenase and catechol dioxygenase genes', *Bioresource technology*, 102(5), pp. 4232-4240.

Tunkel, J., Howard, P.H., Boethling, R.S., Stiteler, W. and Loonen, H. (2000) 'Predicting ready biodegradability in the Japanese Ministry of International Trade and Industry test', *Environmental toxicology and chemistry*, 19(10), pp. 2478-2485.

Tyler, J.E. and Finn, R.K. (1974) 'Growth rates of a pseudomonad on 2, 4-dichlorophenoxyacetic acid and 2, 4-dichlorophenol', *Applied microbiology*, 28(2), pp. 181-184.

UNEP (2013) *Global Chemicals Outlook - Towards Sound Management of Chemicals*. [Online]. Available at: https://sustainabledevelopment.un.org/index.php?page=view&type=400&nr=1966&menu=35.

Upton, M.N., Smith, G.D., McConnachie, A., Hart, C.L. and Watt, G.C.M. (2004) 'Maternal and personal cigarette smoking synergize to increase airflow limitation in adults', *American journal of respiratory and critical care medicine*, 169(4), pp. 479-487.

USGS (2013) 'Emerging contaminants: Transport and fate'. 01/11/2017. Available at: http://toxics.usgs.gov/regional/emc/transport_fate.html

van Beilen, J.B., Wubbolts, M.G. and Witholt, B. (1994) 'Genetics of alkane oxidation byPseudomonas oleovorans', *Biodegradation*, 5(3-4), pp. 161-174.

Van Doorn, R., Klerks, M.M., van Gent-Pelzer, M.P.E., Speksnijder, A., Kowalchuk, G.A. and Schoen, C.D. (2009) 'Accurate quantification of microorganisms in PCR-inhibiting environmental DNA extracts by a novel internal amplification control approach using Biotrove OpenArrays', *Applied and environmental microbiology*, 75(22), pp. 7253-7260.

Vignola, M., Werner, D., Wade, M.J., Meynet, P. and Davenport, R.J. (2017) 'Medium shapes the microbial community of water filters with implications for effluent quality', *Water research*.

Vos, J.G., Dybing, E., Greim, H.A., Ladefoged, O., Lambré, C., Tarazona, J.V., Brandt, I. and Vethaak, A.D. (2000) 'Health effects of endocrine-disrupting chemicals on wildlife, with special reference to the European situation', *Critical reviews in toxicology*, 30(1), pp. 71-133.

Vrede, K., Heldal, M., Norland, S. and Bratbak, G. (2002) 'Elemental composition (C, N, P) and cell volume of exponentially growing and nutrient-limited bacterioplankton', *Applied and Environmental Microbiology*, 68(6), pp. 2965-2971.

Vuono, D.C., Regnery, J., Li, D., Jones, Z.L., Holloway, R.W. and Drewes, J.r.E. (2016) 'rRNA gene expression of abundant and rare activated-sludge microorganisms and growth rate induced micropollutant removal', *Environmental science & technology*, 50(12), pp. 6299-6309.

Wackett, L.P. (2004) 'Stable isotope probing in biodegradation research', *Trends in biotechnology*, 22(4), pp. 153-154.

Wackett, L.P. (2009) 'Questioning our perceptions about evolution of biodegradative enzymes', *Current opinion in microbiology*, 12(3), pp. 244-251.

Wammer, K.H. and Peters, C.A. (2005) 'Polycyclic aromatic hydrocarbon biodegradation rates: a structure-based study', *Environmental science & technology*, 39(8), pp. 2571-2578.

Wang, H., Wang, B., Dong, W. and Hu, X. (2016) 'Co-acclimation of bacterial communities under stresses of hydrocarbons with different structures', *Scientific reports*, 6.

Widada, J., Nojiri, H. and Omori, T. (2002) 'Recent developments in molecular techniques for identification and monitoring of xenobiotic-degrading bacteria and their catabolic genes in bioremediation', *Applied microbiology and biotechnology*, 60(1), pp. 45-59.

Wolfe, N.L., Paris, D.F., Steen, W.C. and Baughman, G.L. (1980) 'Correlation of microbial degradation rates with chemical structure', *Environmental Science & Technology*, 14(9), pp. 1143-1144.

Yang, H., Jiang, Z. and Shi, S. (2006a) 'Aromatic compounds biodegradation under anaerobic conditions and their QSBR models', *Science of the total environment*, 358(1-3), pp. 265-276.

Yang, H., Jiang, Z. and Shi, S. (2006b) 'Aromatic compounds biodegradation under anaerobic conditions and their QSBR models', *Science of the total environment*, 358(1), pp. 265-276.

Yonezawa, Y. and Urushigawa, Y. (1979) 'Chemico-biological interactions in biological purification systems V. Relation between biodegradation rate constants of aliphatic alcohols by activated sludge and their partition coefficients in a 1-octanol-water system', *Chemosphere*, 8(3), pp. 139-142.

Zhang, C. and Hughes, J.B. (2004) 'Bacterial energetics, stoichiometry, and kinetic modeling of 2, 4-Dinitrotoluene biodegradation in a batch respirometer', *Environmental toxicology and chemistry*, 23(12), pp. 2799-2806.

Zhang, Y., Sei, K., Toyama, T., Ike, M., Zhang, J., Yang, M. and Kamagata, Y. (2008) 'Changes of catabolic genes and microbial community structures during biodegradation of nonylphenol ethoxylates and nonylphenol in natural water microcosms', *Biochemical Engineering Journal*, 39(2), pp. 288-296.

Zhou, Y., Huang, H. and Shen, D. (2016) 'Multi-substrate biodegradation interaction of 1, 4-dioxane and BTEX mixtures by Acinetobacter baumannii DD1', *Biodegradation*, 27(1), pp. 37-46.

Zylstra, G.J., McCombie, W.R., Gibson, D.T. and Finette, B.A. (1988) 'Toluene degradation by Pseudomonas putida F1: genetic organization of the tod operon', *Applied and environmental microbiology*, 54(6), pp. 1498-1503.

Appendices

Appendix A

Table A 1 List of Compounds used for generating algorithm to convert ultimate biodegradation rating to biodegradation half-life. B) Summary statistics of regression equation.

A)

Compounds	Ultimate Biodegradation Half-life (days) (eChemPortal)	Ultimate Biodegradation rating (BIOWIN 3)
mannitol	2.33	3.7564
heptanoic acid, methyl ester	8.67	3.319
Quinoline	10	2.914
p-chlorophenol	15	2.7649
napthalene	37.5	2.33
9,10-anthracenedione,1,4-diamino-2,3-dichloro	60	1.7944

B)

	Coefficients	Standard Error	P-value
Intercept	106.05	15.64	0.002
Ultimate Biodegradation rating (BIOWIN 3)	-29.79	5.43	0.005

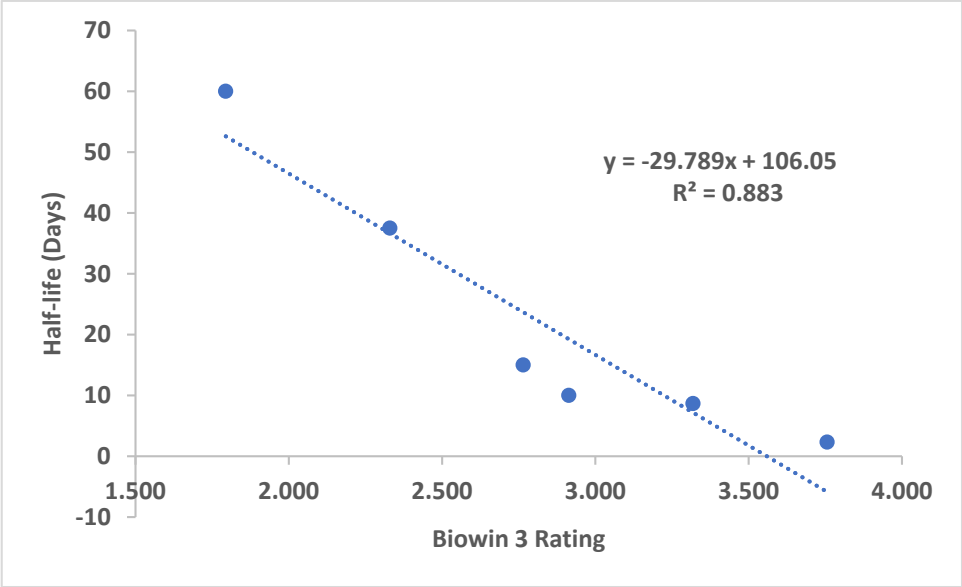


Figure A 1 Statistical regression model to predict the half-life of chemicals

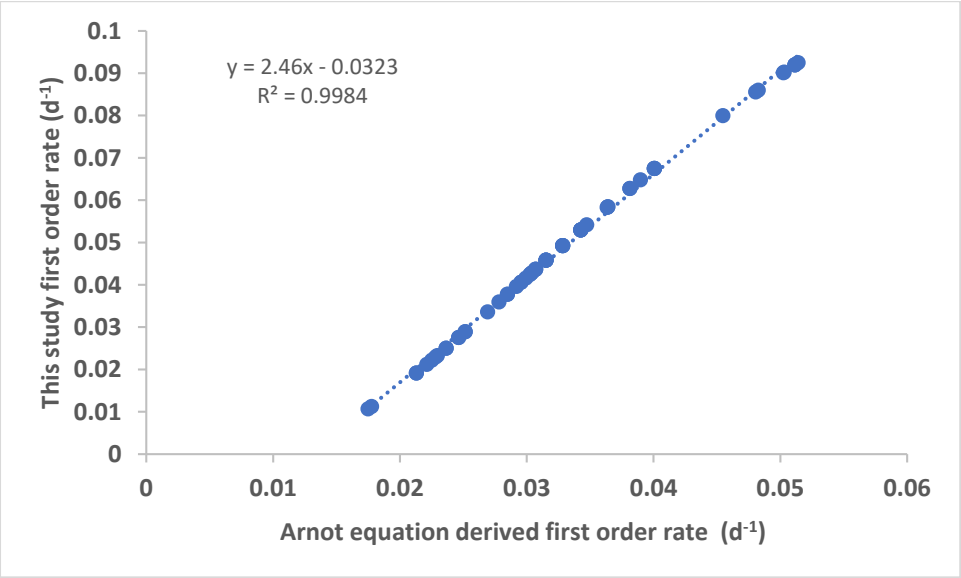


Figure A 2 Relationship between first order rate used in this study and the model developed by (Arnot *et al.*, 2005)

Table A 2 Final list of chemicals (Group 1) used in QSBR model for simple aromatic chemicals

Group 1 (Simple Aromatic Chemicals)					
Prediction Sets		Training Sets			
1	2,4-dibromophenol	1	1,2,4 - Trichlorobenzene	21	3-hydroxy4chlorophenol
2	2,4-dichlorophenol	2	2,4,6-Trichlorophenol	22	nitrobenzene
3	3-iodoaniline	3	2,6-dibromophenol	23	phenylvinylsulfonate
4	4-chloroaniline	4	1,4-Dichlorobenzene	24	p-xylene
5	p-nitrotoluene	5	2-chloro-4-nitrophenol	25	o-xylene
6	1,2,3-trimethylbenzene	6	3-hydroxy-4,6-dichlorophenol	26	2,4-dimethylphenol
7	2-bromophenol	7	4-iodoaniline	27	2,3-dimethylphenol
8	3-methylaniline	8	3-chloroaniline	28	aminobenzoic acid
9	2-chlorophenol	9	4-nitroaniline	29	aminophenol
10	3,4-dimethoxybenzene	10	3-nitroaniline	30	aniline
11	m-xylene	11	4-methylsulfoneaniline	31	4-cynophenol
12	3,4-dimethylphenol	12	1,3,5-trimethylbenzene	32	4-methoxy phenol
13	2,6-dimethylphenol	13	4-iodophenol	33	3-methylphenol
14	isopropylbenzene	14	4-bromophenol	34	2-methylphenol
15	ethylbenzene	15	2-methoxyaniline	35	toluene
16	4-methylphenol	16	nitrobenzoic acid	36	2-carboxylatephenol
17	benzaldehyde	17	aminotoluene	37	Benzoate ion
18	2,5-dihydroxybenzoic acid	18	4-chlorophenol	38	methylbenzoate
19	4-hydroxyphenol	19	chlorobenzene	39	3-hydroxyphenol
		20	1,4-dimethoxybenzene	40	hydroquinone
				41	phenol

Table A 3 Final list of chemicals (Group2) used in QSBR model for complex aromatic chemicals

Group2					
Prediction Sets		**Training Set**			
1	triclosan	1	benzoperylene	10	atenolol
2	chlorobenzilate	2	chlorofenvinphos	11	4chlorobiphenyl
3	levonorgestral	3	ethinylestradiol	12	clofibricacid
4	alachlor	4	progesterone	13	metoprolol
5	estrone	5	bezafibrate	14	carbamazepine
6	17-b-estradiol	6	flumequine	15	ibuprofen
7	bisphenolA	7	diuron	16	4nitrocinnamicacid
8	gemfibrozil	8	sulfamethoxazole	17	4nnonylphenol
9	isoproturon	9	acebutolol	18	octylphenol
10	nonylphenol				

Table A 4 Final list of chemicals (Group3) used in QSBR model for combined aromatic chemicals

Group3					
Prediction Sets		**Training Sets**			
1	benzobfluoranthene	1	benzoperylene	30	gemfibrozil
2	fluoranthene	2	benzoapyrene	31	clofibricacid
3	ethinylestradiol	3	benzofluranthene	32	1,3,5-trimethoxybenzene
4	levonorgestral	4	triclosan	33	isoproturon
5	anthracene	5	chlorofenvinphos	34	bcyclocitral
6	diuron	6	chlorobenzilate	35	1,2,3-trimethylbenzene
7	2,4-dibromophenol	7	progesterone	36	1,3,5-trimethylbenzene

8	2,4-dichlorophenol	8	bezafibrate	37	2-bromophenol
9	3-iodoaniline	9	alachlor	38	4-bromophenol
10	3-chloroaniline	10	flumequine	39	3-methylaniline
11	4-nitroaniline	11	estrone	40	2-chlorophenol
12	atenolol	12	napthalene	41	4-chlorophenol
13	metoprolol	13	2-fluoroaniline	42	3-hydroxy-4-chlorophenol
14	4-iodophenol	14	2,6-dibromophenol	43	1,4-dimethoxybenzene
15	2-methoxyaniline	15	sulfamethoxazole	44	nitrobenzene
16	chlorobenzene	16	benzene	45	m-xylene
17	3,4-dimethoxybenzene	17	17-b-estradiol	46	phenylvinylsulfonate
18	o-xylene	18	3-hydroxy-4,6-dichlorophenol	47	p-xylene
19	2,3-dimethylphenol	19	acebutolol	48	2,4-dimethylphenol
20	3,4-dimethylphenol	20	4-iodoaniline	49	2,6-dimethylphenol
21	4-cynophenol	21	b-ionone	50	acetaminophen
22	anisol	22	4-chloroaniline	51	isopropylbenzene
23	3-methylphenol	23	3-nitroaniline	52	1,4-benzoquinone
24	Benzenesulfonate ion	24	bisphenolA	53	ethylbenzene
25	3-hydroxyphenol	25	4-chlorobiphenyl	54	2-methylphenol
		26	benzaldehyde	55	4-methylphenol
		27	2-carboxylatephenol	56	toluene
		28	Benzoate ion	57	ibuprofen
		29	Methyl benzoate	58	4-hydroxyphenol
				59	phenol

Table A 5 Complete list of 140 chemicals initially selected

Chemicals				
pentabromodiphenylether	sulfamethoxazole	4-iodophenol	ethylebenzene	propene
hexachlorobenzene	benzene	2-bromophenol	anisol	4-methoxycinnamicacid
hexachlorocyclopentadiene	17b-estradiol	4-bromophenol	ketoprofen	3-methoxy-4-hydroxycinnamicacid
trifluralin	2,4-dichlorophenol	cephalexin	2-methylphenol	ethnene
hexachlorocyclohexane	3-hydroxy-4,6-dichlorophenol	2-methoxyaniline	3-methylphenol	vinylacetate
pentachlorobenze	acebutolol	3-methylaniline	4-methylphenol	vinylsulfonateion
medroxyprogesterone	3-iodoaniline	2-chlorophenol	toluene	but-1-en-2-ol
benzoperylene	4-iodoaniline	4-chlorophenol	ibuprofen	diethylhexylphthalate
benzoapyrene	bionone	chlorobenzene	4-nitrocinnamicacid	trans,cis-2,6-nonadienal
benzobfluoranthene	3-chloroaniline	3-hydroxy-4-chlorophenol	vinylphosphonicacid	2-acetamidoacrylicacid
benzofluranthene	4-chloroaniline	1,4-dimethoxybenzene	acrylamide	2-acetamodoacrylicacid
triclosan	3-nitroaniline	3,4-dimethoxybenzene	1-penten-3-one	octadecane
fluoranthene	4-nitroaniline	nitrobenzene	4-n-nonylphenol	cinnamicacid
chlorofenvinphos	bisphenolA	m-xylene	4-nonylphenol	3,4-dihydroxycinnamicacid
chlorobenzilate	4-chlorobiphenyl	o-xylene	acrylonitrile	oseltamiviracid
ethinylestradiol	1,1-dichloropropene	phenylvinylsulfonate	nonylphenol	hexadecane
progesterone	atenolol	p-xylene	vinylbromide	methacrylicacid
levonorgestral	cis-1,2-dichloroethene	anatoxin	vinylcarbonate	acrylicacid
tetrachloroethene	gemfibrozil	diethylvinylphosphonate	vinylenecarbonate	cis-3-hexen-1-ol
bezafibrate	clofibricacid	2,3-dimethylphenol	benzaldehyde	dinbutylphthalate
alachlor	trans-1,2-dichloroethene	2,4-dimethylphenol	benzenesulfonateion	
anthracene	1,1-dichloroethene	2,6-dimethylphenol	tetramethylethene	
flumequine	1,3,5-trimethoxybenzene	3,4-dimethylphenol	octylphenol	
diuron	metoprolol	1,2-dibromoethene	2-carboxylatephenol	

estrone	isoproturon	acetaminophen	benzoateion	
napthalene	b-cyclocitral	isopropylbenzene	methylbenzoate	
trichloroethene	carbamazepine	vinylchloroide	3-hydroxyphenol	
2-fluoroaniline	4-methylsulfoneaniline	4-cynophenol	4-hydroxyphenol	
2,4-dibromophenol	1,2,3-trimethylbenzene	1,4-benzoquinone	phenol	
2,6-dibromophenol	1,3,5-trimethylbenzene	ethylbenzene	2,4,6-trichlorophenol	

Table A 6 List of 60 simple mono-aromatic chemicals, their BIOWIN3 rating, BIOWIN3 derived half-life and the values of different physiochemical and quantum mechanical descriptors for each chemical.

Chemicals	**Status**	**$t_{1/2}$ (d)**	**BIOWIN 3 rating**	**nN**	**Mor08u**	**nArX**	**T.E (eV)**	**E homo (eV)**	**Elomo (eV)**	**E.E (eV)**	**IP**	**MW**	**vdw Volume (Å3)**	**MR (cm^3)**	**Log P**	**(σ)**	**vX1**	**vX2**
2,4-dibromophenol	P	34.2	2.47	0	0.268	2	-1530.82	-9.359	-0.709	-6479.05	9.359	251.90	129.54	45.86	2.68	0.09	3.926	3.366
2,4-dichlorophenol	P	32.5	2.53	0	0.126	2	-1619.95	-9.586	-0.723	-6798.48	9.582	163.00	118.4	37.92	2.88	0.09	3.096	2.439
3-iodoaniline	P	31.0	2.58	1	0.012	1	-1227.45	-8.651	-0.116	-5242.76	8.65	219.03	117.48	43.39	2.07	-0.48	3.378	2.801
3-chloroaniline	P	29.9	2.62	1	-0.018	1	-1271.13	-8.763	-0.087	-5408.1	8.7629	127.57	106.95	35.38	1.75	-0.43	2.677	1.991
3-nitroaniline	P	29.3	2.64	2	-0.085	0	-1777.97	-9.116	-1.159	-7981.6	9.116	138.12	115.97	37.03	1.08	0.12	2.699	1.852
4-methylsulfoneaniline	P	26.7	2.72	1	-0.007	0	-1934.95	-8.999	-0.187	-9605.73	8.998	171.22	145.89	43.81	-0.02	-0.09	4.947	4.357
1,3,5-trimethylbenzene	P	26.1	2.74	0	0.535	0	-1268.04	-9.174	0.535	-6602.52	9.174	120.19	132.57	40.72	3.51	-0.51	3.232	2.665
4-bromophenol	P	25.3	2.77	0	0.083	1	-1321.97	-9.229	-0.371	-5381.64	9.2288	173.01	108.75	35.82	2.44	-0.14	3.027	2.392
3-methylaniline	P	24.7	2.79	1	0.403	0	-1167.7	-8.374	0.464	-5452.04	8.374	107.15	109.87	35.31	1.66	-0.83	2.61	1.914
4-chlorophenol	P	24.4	2.80	0	-0.005	1	-1366.57	-9.336	-0.411	-5498.44	9.336	128.56	104.43	33.02	2.27	-0.14	2.612	1.913
1,4-dimethoxybenzene	P	23.8	2.82	0	0.773	0	-1707.41	-8.378	0.082	-8242.13	8.378	138.16	134.17	39.6	1.66	-0.54	3.053	1.855
m-xylene	P	23.0	2.85	0	0.808	0	-1117.88	-9.228	0.431	-5391.86	9.227	106.17	115.66	35.9	3	-0.34	2.821	2.158
3,4-dimethylphenol	P	22.1	2.88	0	0.738	0	-1413.26	-8.797	0.213	-6811.39	8.797	122.16	124.36	37.78	2.7	-0.71	2.962	2.269
2,4-dimethylphenol	P	22.1	2.88	0	0.488	0	-1413.25	-8.758	0.204	-6816.29	8.758	122.16	124.21	37.78	2.7	-0.71	2.962	2.29
aminophenol	P	21.0	2.92	1	-0.049	0	-1312.88	-8.213	0.031	-5599.32	8.213	109.13	101.45	32.37	0.84	-1.03	2.334	1.592
ethylbenzene	P	20.1	2.95	0	0.227	0	-1117.67	-9.449	0.331	-5393.45	9.449	106.17	115.86	35.8	2.93	-0.15	2.971	1.839
4-methylphenol	P	19.2	2.98	0	0.291	0	-1263.15	-8.927	0.095	-5541.85	8.927	108.14	107.31	32.95	2.18	-0.54	2.545	1.836

toluene	P	19.2	2.98	0	0.355	0	-967.722	-9.426	0.317	-4275	9.426	92.14	98.87	31.07	2.49	-0.17	2.411	1.655
benzoic acid	P	16.3	3.08	0	-0.073	0	-1515.65	-10.446	-2.627	-6307.82	10.445	121.11	107.31	33.18	1.63	0	2.588	1.671
hydroquinone	P	15.5	3.11	0	-0.19	0	-1408.39	-8.796	-0.198	-5703.82	8.796	110.11	99.03	30.01	1.37	-0.74	2.269	1.516
1,2,4- Trichlorobenzene	T	41.2	2.23	0	0.045	3	-1578.02	-9.996	-1.035	-6475.3	9.9964	181.45	123.86	40.93	3.79	0.69	3.439	2.809
2,4,6-Trichlorophenol	T	40.6	2.25	0	0.102	3	-1873.41	-9.639	-1.072	-8052.58	9.63	197.45	132.35	42.81	3.48	0.32	3.579	2.969
2,6-dibromophenol	T	34.2	2.47	0	0.348	2	-1530.75	-9.419	-0.684	-6545.35	9.419	251.90	127.03	43.51	3.21	0.09	3.932	3.279
1,4-Dichlorobenzene	T	33.0	2.51	0	0.117	2	-1324.67	-9.84	-0.679	-5292.42	9.84	147.00	109.85	36.04	3.17	0.46	2.955	2.309
2-chloro-4-nitrophenol	T	32.1	2.54	1	-0.052	1	-2126.94	-10.238	-1.519	-9437.66	10.2375	173.55	127.37	39.57	2.21	0.64	3.117	2.3
3-hydroxy-4,6-dichlorophenol	T	31.9	2.55	0	0.096	2	-1915.35	-9.176	-0.721	-8258.07	9.176	179.00	126.88	39.8	2.57	-0.28	3.236	2.569
4-iodoaniline	T	31.0	2.58	1	0.053	1	-1227.44	-8.334	-0.081	-5231.84	8.334	219.02	117.46	43.39	2.07	-0.48	3.378	2.797
4-chloroaniline	T	29.9	2.62	1	0.19	1	-1271.1	-8.611	-0.107	-5395.43	8.611	127.57	106.95	35.38	1.75	-0.43	2.677	1.988
4-nitroaniline	T	29.3	2.64	2	0.176	0	-1778.13	-9.263	-0.907	-7953.76	9.263	138.12	116	37.03	1.08	0.12	2.699	1.849
p-nitrotoluene	T	27.5	2.70	1	0.47	0	-1728.21	-10.36	-1.184	-7886.1	10.36	137.14	122.8	37.62	2.43	0.61	2.91	2.093
1,2,3-trimethylbenzene	T	26.1	2.74	0	0.722	0	-1267.93	-9.1362	0.515	-6715.65	9.136	120.19	132.83	40.72	3.51	-0.51	3.244	2.514
4-iodophenol	T	25.6	2.76	0	-0.077	1	-1322.9	-8.737	-0.383	-5333.48	8.737	220.01	114.97	41.04	2.6	-0.19	3.313	2.722
2-bromophenol	T	25.3	2.77	0	0.037	1	-1321.91	-9.368	-0.361	-5442.71	9.367	173.01	108.79	35.82	2.44	-0.14	3.033	2.306
2-methoxyaniline	T	25.0	2.78	1	0.278	0	-1462.35	-8.129	0.325	-6947.89	8.129	123.15	119.06	37.16	0.99	-0.93	2.728	1.742
nitrobenzoic acid	T	24.8	2.79	1	-0.228	0	-2291.87	-11.13	-1.697	-10552.6	11.13	167.12	132.71	39.72	1.57	0.78	3.088	2.113
aminotoluene	T	24.6	2.80	1	0.43	0	-1167.61	-8.341	0.406	-5506.97	8.341	107.15	109.9	35.31	1.66	-0.83	2.616	1.857
2-chlorophenol	T	24.4	2.80	0	-0.024	1	-1366.53	-9.453	-0.377	-5567.43	9.4525	128.56	104.48	33.03	2.27	-0.14	2.618	1.859
chlorobenzene	T	24.4	2.80	0	0.091	1	-1071.14	-9.818	-0.249	-4230.06	9.8177	112.55	95.95	31.14	2.58	0.23	2.478	1.732
3,4-dimethoxybenzene	T	23.8	2.82	0	0.287	0	-1707.27	-8.415	0.272	-8416.54	8.4149	138.16	134.17	39.6	1.66	-0.54	3.053	1.855
3-hydroxy4chlorophenol	T	23.8	2.82	0	-0.012	1	-1661.95	-9.126	-0.395	-6942.04	9.126	144.55	112.91	34.91	1.97	-0.51	2.752	2.043
nitrobenzene	T	23.8	2.82	1	0.048	0	-1578.01	-10.726	-1.318	-6650.57	10.7255	123.11	104.98	32.79	1.91	0.78	2.499	1.593
p-xylene	T	23.0	2.85	0	0.77	0	-1117.87	-9.081	0.387	-5379.4	9.081	106.17	115.7	35.9	3	-0.34	2.821	2.155
o-xylene	T	23.0	2.85	0	0.732	0	-1117.79	-9.201	0.401	-5443.26	9.201	106.17	115.84	35.9	3	-0.34	2.827	2.155
2,6-dimethylphenol	T	22.1	2.88	0	0.264	0	-1413.24	-8.867	0.27	-6894.12	8.867	122.16	124.3	37.78	2.7	-0.71	2.968	2.24
2,3-dimethylphenol	T	22.1	2.88	0	0.475	0	-1413.25	-8.953	0.255	-6888.84	8.953	122.16	124.38	37.78	2.7	-0.71	2.962	2.29

aminobenzoic acid	T	21.8	2.89	1	0.228	0	-1731.65	-8.771	-0.5	-8006.23	8.771	137.14	120.86	37.41	1.45	-0.66	2.794	1.892
isopropylbenzene	T	21.0	2.92	0	0.381	0	-1267.62	-9.4366	0.353	-6668.75	9.437	120.19	133.02	40.43	3.22	-0.15	3.354	2.565
aniline	T	20.9	2.92	1	0.016	0	-1017.53	-8.458	0.334	-4332.29	8.458	93.13	93.01	30.48	1.14	-0.66	2.199	1.411
4-cynophenol	T	20.1	2.95	1	-0.16	0	-1408.05	-9.755	-0.765	-5921.48	9.755	119.12	107.07	32.84	1.53	0.29	2.519	1.659
4-methoxy phenol	T	19.7	2.96	0	0.102	0	-1557.9	-8.575	-0.061	-6939.65	8.575	124.14	116.51	34.81	1.51	-0.64	2.657	1.698
3-methylphenol	T	19.2	2.98	0	0.304	0	-1263.19	-9.096	0.164	-5554.96	9.096	108.14	107.31	32.95	2.18	-0.54	2.545	1.839
2-methylphenol	T	19.2	2.98	0	0.211	0	-1263.13	-9.042	0.152	-5615.36	9.042	108.14	107.38	32.95	2.18	-0.54	2.551	1.786
benzaldehyde	T	17.2	3.05	0	0.15	0	-1235.41	-10.169	-0.712	-5133.84	10.169	106.12	101.21	33	1.69	0.42	2.435	1.529
2-carboxylatephenol	T	16.4	3.08	0	-0.139	0	-1812.07	-10.299	-2.984	-7866.51	10.299	138.12	118.34	35.06	1.98	-0.37	2.729	1.822
2,5-dihydroxybenzoic acid	T	15.8	3.10	0	-0.152	0	-2122.4	-9.075	-1.082	-9652.25	9.075	154.12	126.74	36.94	1.67	-0.74	2.863	2.007
methylbenzoate	T	15.8	3.10	0	-0.112	0	-1680.7	-10.339	-0.496	-7930.7	10.339	136.15	127.31	38.08	1.98	-0.17	2.977	1.858
3-hydroxyphenol	T	15.5	3.11	0	-0.185	0	-1408.52	-9.128	-0.032	-5719.89	9.128	110.11	99.01	30.01	1.37	-0.74	2.269	1.52
phenol	T	15.5	3.11	0	-0.113	0	-1113.02	-9.235	0.036	-4432.63	9.235	94.11	90.52	28.13	1.67	-0.37	2.134	1.336
cathecol	T	15.5	3.11	0	0.174	0	-1407.25	-9.068	-0.084	-5735.71	9.068	110.11	99.01	30.01	1.37	-0.74	2.275	1.489
phenylvinylsulfonate	T	23.0	2.85	0	0.204	0	-2152.26	-9.361	-0.449	-10776.6	9.36	184.23	153.54	46.33	1.81	0.35	4.364	3.242

nN : Number of nitrogen atom (Constitutional Index), **Mor08u** : un-weighted descriptor with scattering parameter (s) = 7 $Å^{-1}$ (3D Molecular Representations of Structure based on Electron diffraction), **nArX** : Number of halogen on aromatic rings (Functional group counts), $\mathbf{E_{homo}}$: Energy of highest occupied molecular orbital (Quantum-mechanical**),** $\mathbf{E_{lomo}}$: Energy of lowest occupied molecular orbital (Quantum-mechanical), **E.E** : Electrical Energy (Quantum-mechanical), **T.E**: Total Energy (Quantum-mechanical), **MW**: Molecular Weight (Size), **vdw Volume**: van der wal volume (Steric), **MR**: Molar Refractivity (Steric), **Log P**: Octanol water partitioning coefficient (Lipophilic), **σ** : Substituent constant (electronic), **vX1 & vX2** : valence connectivity indices (Steric)

P: Prediction

T: Training

$t_{1/2}$: half-life

1. **Goodness of Fit**

$$R^2 = 1 - \frac{\sum_{i=1}^{n}(y_i - \hat{y}_i)^2}{\sum_{i=1}^{n}(y_i - \bar{y})^2} \quad (1)$$

$$RMSEC = \sqrt{\frac{\sum_{i=1}^{n}((y_i - \hat{y}_i)^2}{n}} \quad (2)$$

where y_i – experimental (observed) value of the endpoint for the i[th] compound from the training set; $\hat{y}_i$– predicted value of the endpoint for the i[th] compound from the training set; $\bar{y}$ – the mean experimental value of the endpoint in the training set; n – the number of compounds in the training set.

2. **Robustness**

$$Q^2_{LOO} = 1 - \frac{\sum_{i=1}^{n}(y_i - \hat{y}_{i/i})^2}{\sum_{i=1}^{n}(y_i - \bar{y})^2} \quad (3)$$

$$RMSECV = \sqrt{\frac{\sum_{i=1}^{n}(y_i - \hat{y}_{i/i})^2}{n}} \quad (4)$$

where y_i – experimental (observed) value of the endpoint for the i[th] compound from the training set; $\hat{y}_{i/i}$ – response of the *i-th* object estimated by using a model obtained without using the *i-th object*; $\bar{y}$ – the mean experimental value of the endpoint in the training set; n – the number of compounds in the training set.

3. **Predictive ability**

$$Q^2_{F1} = 1 - \frac{\sum_{i=1}^{nEXT}(y_i - \hat{y}_i)^2}{\sum_{i=1}^{nEXT}(\hat{y}_i - \bar{y}_{TR})^2} = 1 - \frac{PRESS}{SS_{EXT}(\bar{y}_{TR})} \quad (5)$$

$$Q^2_{F2} = 1 - \frac{\sum_{i=1}^{nEXT}(y_i - \hat{y}_i)^2}{\sum_{i=1}^{nEXT}(\hat{y}_i - \bar{y}_{EXT})^2} = 1 - \frac{PRESS}{SS_{EXT}(\bar{y}_{EXT})} \quad (6)$$

$$Q^2_{F3} = 1 - \frac{[\sum_{i=1}^{nEXT}(y_i - \hat{y}_i)^2]/n^{EXT}}{[\sum_{i=1}^{nTR}(\hat{y}_i - \bar{y}_{TR})^2]/n^{TR}} = 1 - \frac{PRESS/n^{EXT}}{TSS/n^{TR}} \quad (7)$$

$$RMSEP = \sqrt{\frac{\sum_{i=1}^{nEXT}(y_i - \hat{y}_i)^2}{n^{EXT}}} \quad (8)$$

where y_i and $\hat{y}_i$ are predicted and observed value of endpoints for the i[th] compound from the validation set respectively; $\bar{y}_{EXT}$ is the response means of the prediction set; $\bar{y}_{TR}$ is the response mean of training set; PRESS is the predictive sum of squares; $SS_{EXT}(\bar{y}_{EXT})$ the total sum of squares of the external set calculated by means of external set mean; TSS is the total sum of squares; n^{TR} is the number of compounds in the training set; n^{EXT} is the number of compounds in the validation set.

4. **Leverage Value**

$$h_i = x_i^T(X^TX)^{-1}x_i \quad (9)$$

where, x_i is A row vector of molecular descriptors for particular (i-th) compound, X is a matrix of descriptors for the training set, X^T is the transpose of matrix of descriptors, $(X^TX)^{-1}$

is the inverted matrix based on descriptors X for all training compounds and x_i^T is the transpose of a row vector of molecular descriptors for particular (i-th) compound.

5. 3D-MoRSE of Aromatic compounds

3D MORSE denotes 3D Molecular Representations of Structure based on Electron diffraction and is defined as:

$$\mathbf{Mor}(\mathbf{s},\mathbf{r}) = \mathbf{I}(\mathbf{s},\mathbf{r}) = \sum_{i=2}^{N}\sum_{j=1}^{i-1} A_j A_i \frac{\sin s\, r_{ij}}{s\, r_{ij}} \quad (10)$$

Where, s is the scattering parameter (s value ranges from 0 to 31 Å^{-1}), $\boldsymbol{r_{ij}}$ is the Euclidean distance between the i^{th} and j^{th} atoms, N is the total number of atoms, A_j and A_i are different atomic properties used as weights (u is un-weighted, m is weighted by mass, v is weighted by van der Waals volume, e is weighted by electronegativity and p is weighted by polarizability). The 3D- MoRSE descriptor used in the above model (i.e. Mor08u) is an un-weighted descriptor with scattering parameter (s) = 7 Å^{-1}.

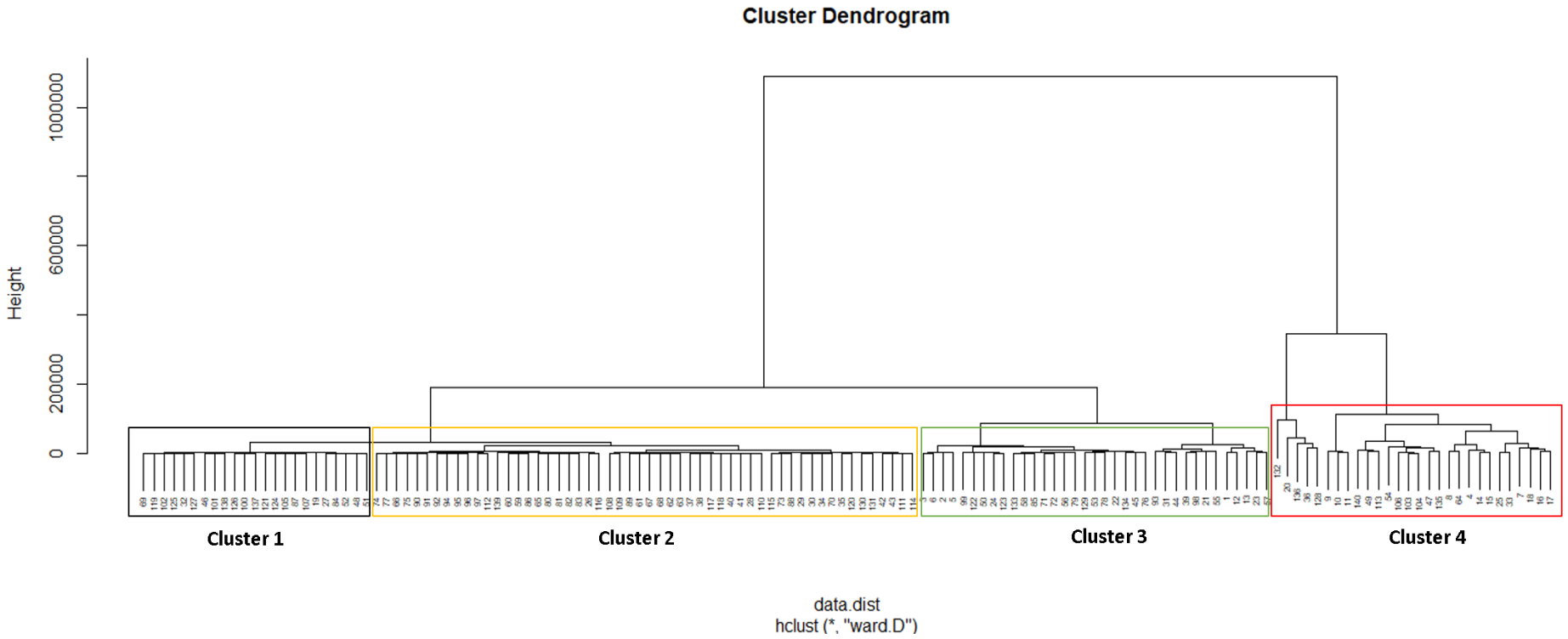


Figure A 3 Hierarchical cluster Analysis to recognize the pattern of 141 chemicals based on Dragon descriptors used for QSBR modelling.

Appendix B

1. Composition of Trace element solution

Trace element solution SL-4:

EDTA	0.50 g
$FeSO_4$ x 7 H_2O	0.20 g
Trace element solution SL-6 (see below)	100.00 ml
Distilled water	900.00 ml

Trace element solution SL-6:

$ZnSO_4$ x 7 H_2O	0.10 g
$MnCl_2$ x 4 H_2O	0.03 g
H_3BO_3	0.30 g
$CoCl_2$ x 6 H_2O	0.20 g
$CuCl_2$ x 2 H_2O	0.01 g
$NiCl_2$ x 6 H_2O	0.02 g
Na_2MoO4 x 2 H_2O	0.03 g
Distilled water	1000.00 ml

2. Analysis of chemical removal; HPLC.

The removal of studied chemicals in the bioreactors over the duration of experiment were quantified in HPLC equipped with UV detector. 1mL of sample were clarified by centrifugation (8000 xg, 5 min) followed by filtration (through 13 mm diameter, 0.2 µm PVDF syringe filter). 0.5 mL of filtrate was mixed with equivalent volume of methanol and analysed to quantify the chemical removal in HPLC. **Table.7** shows the detailed method and programs used for quantification of individual chemicals.

Table B 1 Details of method to analyse chemicals in HPLC

Chemical	Technique	Column	Solvent	Flow rate ($mL.min^{-1}$)	Detector	Injection volume (uL)
Phenol	HPLC	NX-C18 column (100 ✕ 4.6 mm; 3 um particle size; Phenomen ex)	Methanol:Water (50:50)	1	UV/Visible detector at 280 nm	20
4-CP	HPLC	150x4.6m m 5µm Fluophase PFP HPLC Column (Thermo Scientific)	Methanol:Water (50:50)	1	UV/Visible detector at 280 nm	20
2,4-DCP	HPLC	150x4.6m m 5µm Fluophase PFP HPLC Column (Thermo Scientific)	Methanol:Water (50:50)	1	UV/Visible detector at 280 nm	20
m-cresol	HPLC	NX-C18 column (100 ✕ 4.6 mm; 3 um particle size; Phenomen ex)	Methanol:Water (50:50)	1	UV/Visible detector at 280 nm	20

3. GC-FID analysis, toluene (headspace and aqueous phase) analysis

The headspace concentration of toluene and aqueous phase toluene in the bioreactors were measured using a HP-7890A Series Gas Chromatograph (Agilent Technologies, Palo Alto, USA). Briefly, for headspace analysis, 100 µL of gas was injected manually using a Hamilton

gas-tight syringe into the machine in split mode, the injector set at 200 °C, flame ionisation detector at 300 °C. Prior to this, a blank sample containing air was ran to ensure that there had not been any previous contamination of the GC columns. This was followed by a calibration of the instrument using different volumes of the toluene. The standard was prepared by adding the known volume of toluene in clean closed glass bottles at room temperature and left to evaporate overnight. The fraction of chemicals evaporated were calculated using ideal gas law. The separation of headspace gas was performed on a fused silica capillary column (30 m x 0.25 mm i.d) coated with 0.25µm dimethyl poly-siloxane (HP-5 phase). To analyse toluene, the column temperature was maintained at 30 °C for 10 minutes and raised to 180 °C at a rate of 15 °C min^{-1} and then held at this temperature for 20 minutes.

The toluene in the aqueous phase was extracted in hexane using liquid- liquid extraction technique and hexane extract was analysed in HP-7890A Series Gas Chromatograph (Agilent Technologies, Palo Alto, USA). Briefly 1ml of aqueous sample was mixed with 2 mL of hexane in 10 mL glass tube with screw cap and mixed briefly my inverting the tube gently. The hexane layer was withdrawn using glass pipette and transferred into new 5 ml glass tube. Residual aqueous media was removed using anhydrous sodium sulfate. The hexane was blown with nitrogen gas to reduce the final volume to 1 mL. The hexane extract was analysed in HP-7890A Series Gas Chromatograph equipped with flame ionization detector. The separation was performed on fused silica capillary column (30 m x 0.25 mm i.d) coated with 0.25µm dimethyl poly-siloxane (HP-5 phase). The column temperature was maintained at 30 °C for 10 minutes and raised to 180 °C at a rate of 15 °C min^{-1} and then held at thistemperature for 20 minutes. The standard solution of toluene was used to prepare calibration series in hexane ranging between 0.5 mg/L to 100 mg/L. The GC data was acquired on a Thermo-Atlas laboratory data system on channel 2.

4. Flow cytometer; quantification of microbial growth

Fluorescent staining of Bacteria

To distinguish viable and dead bacteria, the bacteria sample were stained with SYBR Green I (1:1000 dilution of commercial stock; Invitrogen, USA; λ_{ex} = 495 nm, λ_{em} = 525 nm) diluted in dimethyl sulfoxide (DMSO; Merck, Germany) and 5 µL.mL^{-1} Propidium Iodide (PI, stock solution concentration 1mg mL-1; Invitrogen, USA; λ_{ex} = 536 nm, λ_{em} = 617 nm). Samples from bioreactor were diluted with sterile Phosphate Buffered Saline (PBS, 8 g/L NaCl, 0.2 g/L KCl, 1.15 g/L Na_2HPO_4, 0.2 g/L KH_2PO_4; pH 7.3) to achieve an event rate between 200 and 800 bacteria s^{-1} and avoid coincidence (i.e. two or more bacteria and/or particles being simultaneously within the sensing zone). Diluted samples were then stained with SYBR I and PI using the protocol of (Hammes *et al.*, 2008) with some modification. Briefly, bacterial cells

were stained with 10 µL.mL^{-1} SYBR Green I and 5 µL.mL^{-1} Propidium Iodide, incubated in the dark at 37ºC for at least 15 min just before measurement in FACScan flow cytometer. SYBR-I is capable of staining all cells, whereas the polarity of PI allows it to penetrate only the cells with permeabilised membranes, characteristics of dead cells (Foladori *et al.*, 2010). In dead cells, simultaneous staining with SYBR I and PI activates energy transfer between the fluorochromes. As, a consequence, viable bacteria emit green fluorescence, while dead bacteria emit red fluorescence.

Flow Cytometer

The growth of degraders in the bioreactor over the period of experiment was quantified using FACScan flow cytometer (Becton Dickinson, California) equipped with a 15-mW 488-nm air-cooled argon-ion laser and a standard filter setup. Data acquisition was gated on the green fluorescent distribution in order to acquire the signals of viable or active bacteria only, and to eliminate scattering produced by non-bacterial particles, non-viable or inactive bacteria and other derbies. Green and red fluorescence were collected with logarithmic gain.

5. CO_2 analysis by GC-MS analysis

In order to monitor the mineralization process of each chemical in the bioreactor over the duration of the experiments, the development of CO2 in the headspace of each bioreactor was measured using Gas Chromatography. Briefly, the quantification of headspace CO_2 was conducted on a Fisons 8060 GC using spilt injection (150 °C) linked to a Fisons MD800MS (electron voltage 70eV, filament current 4A, source current 800 µA, source temperature 200 °C, multiplier voltage 500 V, interface temperature 150 °C). The acquisition of data was controlled by a Compaq Deskpro computer using the Xcalibur software in full scan mode (1-151 amu/sec dwell 10ms/amu). A series of CO_2 standard and headspace sample (100 uL) was injected in split mode and the GC programme and MS data acquisition commenced. Separation was performed on a HP-PLOT-Q capillary column (30 mm x 0.32 mm i.d) packed with 20 µm Q phase. The GC was held isothermally at 35 °C with helium as the carrier gas (flow rate 1 mL/min, pressure of 60 kPa, open split at 100 mls/min). The chromatograms of the standard and headspace gas (CO_2) were integrated and quantified and the gas concentrations deduced. Calculations were done based on the assumptions of the ideal gas law using the following conditions: temperature = 293.15K, atmospheric pressure = 1 atm, volume = 1L and a gas constant R = 0.0821 L atm K^{-1}mol^{-1}.

6. Batch biodegradation model

The batch experiments were simulated by assuming Monod kinetic growth of microbial biomass following the addition of substrates (Phenol, m-cresol, toluene, 4-chlorophenol and 2,4-dichlorophenol), limited by a maximum biomass concentration in the aqueous medium. Parameters are

expressed in SI units of mole, second, kg, meters, except for the hydrogen ion concentration [H^+] which is expressed in moles per litre. Substrate, biomass and CO_2 concentrations are expressed on a carbon-normalized basis (mole substrate C per m^3, mole biomass C per m^3, mole CO_2 C per m^3) to facilitate elemental mass balances.

Table B 2 Independent and dependent variables, parameters and their dimensions

$C_{air}^{CO_2}$ (moles C m^{-3})	CO_2 concentration in bioreactor headspace
$C_{water}^{CO_2}$ (moles C m^{-3})	CO_2 concentration in bioreactor aqueous medium
H^{CO_2} (-)	Dimensionless Henry constant for CO_2
$K_h^{CO_2}$ (-)	Hydration constant for CO_2
$C_{water}^{H_2CO_3^*}$ (moles C m^{-3})	Apparent carbonic acid concentration in water
$C_{water}^{H_2CO_3}$ (moles C m^{-3})	Carbonic acid concentration in water
$H_*^{CO_2}$(-)	Apparent dimensionless Henry constant for CO_2
K_1 (moles per litre)	Apparent carbonic acid dissociation constant
K_2 (moles per litre)	Hydrogen carbonate dissociation constant
$C_{water}^{HCO_3^-}$ (moles C m^{-3})	Hydrogen carbonate concentration in water
H^+ (moles per litre)	Hydrogen ion concentration in water
$C_{water}^{CO_3^{2-}}$(moles C m^{-3})	Carbonate concentration in water
$N_{tot}^{CO_2}$ (moles C)	Total amount of CO_2 in the bioreactor
V_{water} (m^3)	Volume of aqueous medium in batch
μ_{max}^{sub} (s^{-1})	Maximum specific biomass growth rate
C_{water}^{sub} (moles C m^{-3})	Substrate concentration in aqueous medium
K_S^{sub} (moles C m^{-3})	The half-velocity constant
Y^{sub}(moles C $moles^{-1}$ C)	Yield coefficient
C_{water}^{bio}(moles C m^{-3})	Biomass concentration in aqueous medium
V_{air} (m^3)	Volume of headspace in bioreactor
H^{sub} (-)	Dimensionless Henry constant for the substrate
t (s)	Time

CO_2

The CO_2 concentration in bioreactor headspace air, $C_{air}^{CO_2}$, is related to the dissolved CO_2 concentration in aqueous medium, $C_{water}^{CO_2}$, by Henry's law

$$H^{CO_2} = \frac{C_{air}^{CO_2}}{C_{water}^{CO_2}} \quad (6.1)$$

The equilibrium between dissolved CO_2 and carbonic acid (H_2CO_3) is described by a hydration equilibrium constant, $K_h^{CO_2}$,

$$K_h^{CO_2} = \frac{C_{water}^{H_2CO_3}}{C_{water}^{CO_2}} \tag{6.2}$$

The sum of dissolved CO_2 and H_2CO_3 concentration will be referred as the apparent carbonic acid concentration in water, $C_{water}^{H_2CO_3^*}$,

$$C_{water}^{H_2CO_3^*} = C_{water}^{CO_2} + C_{water}^{H_2CO_3} \tag{6.3}$$

And the modified Henry's law constant relates the CO_2 concentration in the bioreactor headspace air to the apparent carbonic acid concentration in the aqueous medium

$$H_*^{CO_2} = \frac{C_{air}^{CO_2}}{C_{water}^{H_2CO_3^*}} \tag{6.4}$$

The apparent carbonic acid dissociation constant K_1, relates the apparent carbonic acid concentration in the aqueous medium to the hydrogen carbonate (or bicarbonate) concentration in the aqueous medium, $C_{water}^{HCO_3^-}$,

$$K_1 = \frac{C_{water}^{HCO_3^-} \cdot [H^+]}{C_{water}^{H_2CO_3^*}} \tag{6.5}$$

where $[H^+]$ is the molar hydrogen ion concentration, or 10^{-pH}.

The hydrogen carbonate dissociation constant, K_2, relates the hydrogen carbonate concentration in the aqueous medium to the carbonate concentration in the aqueous medium, $C_{water}^{CO_3^{2-}}$,

$$K_2 = \frac{C_{water}^{CO_3^{2-}} [H^+]}{C_{water}^{HCO_3^-}} \tag{6.6}$$

The total amount of CO_2 in the system, $N_{tot}^{CO_2}$, is given by

$$N_{tot}^{CO_2} = (V_{air} + V_{water} \cdot \frac{1}{H_*^{CO2}} \left(1 + \frac{K_1}{10^{-pH}} + \frac{K_1 K_2}{10^{-2pH}}\right)) \cdot C_{air}^{CO_2} \tag{6.7}$$

Substrate

The dimensionless Henry constant relates the substrate concentration in the bioreactor headspace air to the apparent substrate concentration in the aqueous medium

$$H^{sub} = \frac{C_{air}^{sub}}{C_{water}^{sub}} \quad (6.8)$$

The substrate mass utilization rate due to biodegradation, r_1, is described by assuming Monod kinetics limited by logistic biomass growth according to

$$(V_{water} + V_{air} \cdot H^{sub}) \, {}^{d}/_{dt} \, C_{water}^{sub} = -r_1 \quad (6.9)$$

$$r_1 = V_{water} \cdot \frac{1}{Y^{sub}} \cdot \mu_{max}^{sub} \cdot \frac{C_{water}^{sub}}{K_S^{sub} + C_{water}^{sub}} \cdot C_{water}^{bio} \quad (6.10)$$

or 0, if t < lag phase.

μ_{max}^{sub} is the maximum specific growth rate, K_S^{sub} is the half-velocity constant, Y^{sub} is the yield coefficient, and C_{water}^{bio} is the concentration of the biomass in the aqueous medium.

Biomass

The following differential equation governs the growth of biomass on the substrate

$$V_{water} \, {}^{d}/_{dt} \, C_{water}^{bio} = r_2 \quad (6.11)$$

$$r_2 = V_{water} \cdot \mu_{max}^{sub} \cdot \frac{C_{water}^{sub}}{K_S^{sub} + C_{water}^{sub}} \cdot C_{water}^{bio} \quad (6.12)$$

Or 0, if t < lag phase.

Assuming instantaneous exchange of CO_2 between headspace air and aqueous medium, a constant aqueous medium pH, and no carbonate dissolution or precipitation, the following differential equations govern the concentration of CO_2 bioreactor headspace air.

$$\left(V_{air} + V_{water} \cdot \frac{1}{H_*^{CO2}} \left(1 + \frac{K_1}{10^{-pH}} + \frac{K_1 K_2}{10^{-2pH}}\right)\right) {}^{d}/_{dt} \, C_{air}^{CO_2} = r_3 \quad (6.13)$$

the net CO_2 mass production rate, r_3, is described by

$$r_3 = V_{water} \cdot \mu_{max}^{sub} \cdot \frac{C_{water}^{sub}}{K_S^{sub} + C_{water}^{sub}} \cdot \frac{1 - Y^{sub}}{Y^{sub}} \cdot C_{water}^{bio} \quad (6.14)$$

Or 0, if t < lag phase.

Numerical solution

The system of ordinary differential equations is solved with Matlab © using the ordinary differential equation solver ode15. The system of ordinary differential equations is

$$d/_{dt}\, C_{water}^{sub} = -\frac{r_1}{(V_{water}+V_{air}\cdot H^{sub})} \quad (6.15)$$

$$d/_{dt}\, C_{water}^{bio} = \frac{r_2}{V_{water}} \quad (6.16)$$

$$d/_{dt}\, C_{air}^{CO_2} = \frac{r_3}{(V_{air}+V_{water}\cdot\frac{1}{H_*^{CO2}}\left(1+\frac{K_1}{10^{-pH}}+\frac{K_1K_2}{10^{-2pH}}\right))} \quad (6.17)$$

The initial conditions are defined by the substrate and biomass concentration in the aqueous medium, and the CO_2 concentration in the headspace at time zero.

Parameter fitting procedure

The lag phase was determined based on the time required for 10 % removal of test chemical, where applicable. The yield coefficient Y^{sub} was determined from the measure total amounts of CO_2 carbon and biomass carbon in the bioreactor at the end of the experiments, when the test chemical (carbon substrate) had been completely degraded. For the remaining Monod kinetic parameters, μ_{max}^{sub} and K_S^{sub}, a two parameter fitting procedure was developed. After defining a plausible range for the values of each parameter, ten parameter with equal increments from the minimum to the maximum were then defined to cover this range. The model was then repeatedly run to test the agreement with the measurements for all the 100 possible combinations of the predefined values of the two parameters. The best μ_{max}^{sub} and K_S^{sub} parameter value combination was identified as that having the least sum of squared residuals for the agreement between modelled and measured data, considering all the available measurements (chemical, CO_2 and biomass concentrations). A further optimization was then conducted after narrowing the range of the parameter values around the previously identified optimum. It should be noted that a fixed ratio of μ_{max}^{sub} to K_S^{sub} tends to give very similar agreements between measured and modelled data which is a known issue for the Monod kinetic model, and implies that $\mu_{max}^{sub}/K_S^{sub}$ is generally more robust than either of the two parameters (Healey, 1980b; Kovárová-Kovar and Egli, 1998).

Biomass carbon content

Biomass carbon content/m^3 = carbon content per bacteria cell *

total bacteria count/m^3 (6.18)

7. Batch Model Script (for 4-chlorophenol) in MatLab

```
clear all
close all
clc

% Simulates a kinetic batch biodegradation test with Monod kinetic growth
% of pollutant degrading biomass on a carbon substrate limited by logistic
% growth (i.e. a maximum biomass carrying capacity).
```

```
% The model links substrate consumption, biomass growth and CO2 production
% with a carbon mass balance.
%
% t: Time [s]
% c: Vector of dependant variables
% 1: Substrate concentration in batch water [moles substrate C/m3]
% 2: Biomass concentration in batch water [moles biomass C/m3]
% 3: CO2 concentration in headspace air [moles CO2 C/m3]


%% Set parameters


% Batch
% Total volume batch, Vbatch [m3]
Vbatch = 580/10^6;


% Water
% Batch water volume [m3]
Vw = 200/1000000;
% pH
pH = 7.053;


% Substrate (i.e. pollutant)
% Dimensionless Henry coefficient, H [moles substrate C per m3 air/moles substrate C per m3 water]
H = 2.29*10^-5;


% Biomass and biodegradation
% Fitting range for Monod half-rate constant, KS [moles substrate C/m3]
KS_min = 0;
KS_max = 7.5;
KS_vector=[0:((KS_max-KS_min)/9):(KS_max-KS_min)]'+KS_min*ones(10,1);
KS_range1 = KS_vector;
% Fitting range Monod maximum growth rate, umax [1/s]
umax_min = 0;
umax_max = 0.001;
umax_vector=[0:((umax_max-umax_min)/9):(umax_max-umax_min)]'+umax_min*ones(10,1);
umax_range1 = umax_vector;
%   Yield coefficient, Y [moles biomass C/moles substrate C]
Y = 0.352;
%   Lag phase, lag [s]
lag = 5*3600;
%   Cell carbon content [moles biomass C/cell]
```

```
Ccell = 9.4*10^-14/12;

% CO2
% Initial concentration [moles CO2-C/m3]
C_CO2_0 = 0.045;
% Acid constants CO2 [moles/L]
K1 = 4.45*10^-7;
K2 = 4.69*10^-11;
% Dimensionless Henry constant air-water CO2 [moles CO2 C per m3 air/moles CO2 C per m3 water]
H_CO2 = 0.034/0.0404;

%% Calculated parameters

% Batch air volume [m3]
Va = Vbatch-Vw;

%% Initial conditions and duration

% Amount of substrate added to the batch [moles substrate C]
MP0 = 0.000467;
% Initial substrate concentration in batch water [moles/m3]
Cw0 = MP0/(Vw+Va*H);
% Initial biomass concentration [moles biomass C/m3 of water]
Cb0 = 3.33*10^13*Ccell;
c0 = zeros(3,1);
c0(1,1)=Cw0;
c0(2,1)=Cb0;
c0(3,1)=C_CO2_0;

% Duration of the experiment [s]
duration = 2*24*60*60;
tspan = [0 duration];

%% Experimental data
% Substrate data [time in s Conc in moles C/m3]
Sub_data = [0*3600 2.333;...
    4.75*3600 2.117;...
    7.5*3600 2.160;...
    17.5*3600 2.022;...
    20*3600 1.870;...
```

```
    23.16*3600 1.727;...
    26*3600 1.448;...
    28.5*3600 0.900;...
    32.5*3600 0.245;...
    40.75*3600 0.0312];

% CO2 data [time in s Conc in moles CO2 C/m3]
CO2_data = [0*3600 0.045;...
    4.75*3600 0.07136;...
    7.5*3600 0.1131;...
    17.5*3600 0.1213;...
    20*3600 0.1120;...
    23.16*3600 0.1390;...
    26*3600 0.1520;...
    28.5*3600 0.2193;...
    40.75*3600 0.2388];

% Cell data [time in s Biomass C in moles/m3]
Cell_data = [0*3600 0.2779;...
    4.75*3600 0.2832;...
    7.5*3600 0.3307;...
    17.5*3600 0.3796;...
    20*3600 0.4032;...
    23.16*3600 0.6108;...
    26*3600 0.8159;...
    28.5*3600 0.8381;...
    40.75*3600 1.0511];

%% Calculate areas, volumes and capacities
% Radii of porous particle shells, including the innermost radius of zero

% Compute the capacity for each element [m3] or [kg]
Capacity=[Vw+Va*H;Vw;Va+Vw/H_CO2*(1+K1/10^-pH+K1*K2/10^-(2*pH))];

%% Determine sparsity pattern
ivec=nan(3+4,1);
jvec=nan(3+4,1);
avec=ones(3+4,1);
% self relationship
index = 0;
add = 3;
```

```
ivec(index+1:index+add)=1:3;
jvec(index+1:index+add)=1:3;
index = index+add;
% Substrate in batch water -> biomass relationship
add = 1;
ivec(index+1:index+add)=1;
jvec(index+1:index+add)=2;
index = index+add;
% Biomass -> substrate in batch water relationship
add = 1;
ivec(index+1:index+add)=2;
jvec(index+1:index+add)=1;
index = index+add;
% CO2 -> substrate in batch water relationship
add = 1;
ivec(index+1:index+add)=3;
jvec(index+1:index+add)=1;
index = index+add;
% CO2 -> biomass in batch water relationship
add = 1;
ivec(index+1:index+add)=3;
jvec(index+1:index+add)=2;
index = index+add;
% Make a sparse matrix of that
sparsepat=sparse(ivec,jvec,avec);

%% Storage of SumSqResiduals for parameter fitting
SumSquaredResiduals1 = zeros(10,10);
SumSquaredResiduals2 = zeros(10,10);
SumSqResidualsStore = zeros(10,10);

%% Set options (absolute and relative tolerance, sparsity pattern,
%              enforce non-negativity for all concentrations)
options=odeset('abstol',1e-13,'reltol',1e-7,'Jpattern',sparsepat,...
            'nonnegative',[1:3]);

%% Iteration for parameter optimization
for k=1:2
minimum = 10^12
for x=1:10
for y=1:10
```

```
KS = KS_vector(x);
umax = umax_vector(y);
%% Solving the system of differential equations
[time,concentrations] = ode15s(@BatchDGL_Kishor,tspan,c0,options,...
   Vw,lag,umax,KS,Y,Capacity);

%% Extracting solutions and plotting
% Substrate prediction water [moles C/m3]
Cw=concentrations(:,1);
% Biomass prediction [moles C/m3]
Cb=concentrations(:,2);
% CO2 prediction concentration in air [moles C/m3]
CO2=concentrations(:,3);

PredictionsSubstrate = interp1(time,Cw,Sub_data(:,1));
ResidualsSubstrate = PredictionsSubstrate-Sub_data(:,2);
SquaredResidualsSubstrate = ResidualsSubstrate.*ResidualsSubstrate;

PredictionsCO2 = interp1(time,CO2,CO2_data(:,1));
ResidualsCO2 = PredictionsCO2-CO2_data(:,2);
SquaredResidualsCO2 = ResidualsCO2.*ResidualsCO2;

PredictionsCell = interp1(time,Cb,Cell_data(:,1));
ResidualsCell = PredictionsCell-Cell_data(:,2);
SquaredResidualsCell = ResidualsCell.*ResidualsCell;

SumSquaredResiduals =
sum(SquaredResidualsSubstrate)+sum(SquaredResidualsCO2)+sum(SquaredResidualsCell);

SumSqResidualsStore(x,y) = SumSquaredResiduals;

if SumSquaredResiduals < minimum
minimum = SumSquaredResiduals;
KS_bestfit = KS
umax_bestfit = umax
end

end
end
```

```
if k == 1
   SumSquaredResiduals1 = SumSqResidualsStore;
end
if k == 2
   SumSquaredResiduals2 = SumSqResidualsStore;
end

if k == 1
% Defining narrower parameter value range around previous optimum
KS_min_new = max(KS_min,KS_bestfit-(KS_max-KS_min)/9);
KS_max_new = min(KS_max,KS_bestfit+(KS_max-KS_min)/9);
umax_min_new = max(umax_min,umax_bestfit-(umax_max-umax_min)/9);
umax_max_new = min(umax_max,umax_bestfit+(umax_max-umax_min)/9);
KS_vector=[0:((KS_max_new-KS_min_new)/9):(KS_max_new-KS_min_new)]'+KS_min_new*ones(10,1);
umax_vector=[0:((umax_max_new-umax_min_new)/9):(umax_max_new-
umax_min_new)]'+umax_min_new*ones(10,1);
KS_range2 = KS_vector;
umax_range2 = umax_vector;
end

end

% Run with fitted parameter values
KS = KS_bestfit;
umax = umax_bestfit;

%% Solving the system of differential equations
[time,concentrations] = ode15s(@BatchDGL_Kishor,tspan,c0,options,...
   Vw,lag,umax,KS,Y,Capacity);

%% Extracting solutions and plotting
% Substrate prediction water [moles C/m3]
Cw=concentrations(:,1);
% Biomass prediction [moles C/m3]
Cb=concentrations(:,2);
% CO2 prediction concentration in air [moles C/m3]
CO2=concentrations(:,3);

PredictionsSubstrate = interp1(time,Cw,Sub_data(:,1));
ResidualsSubstrate = PredictionsSubstrate-Sub_data(:,2);
SquaredResidualsSubstrate = ResidualsSubstrate.*ResidualsSubstrate;
```

```
PredictionsCO2 = interp1(time,CO2,CO2_data(:,1));
ResidualsCO2 = PredictionsCO2-CO2_data(:,2);
SquaredResidualsCO2 = ResidualsCO2.*ResidualsCO2;

PredictionsCell = interp1(time,Cb,Cell_data(:,1));
ResidualsCell = PredictionsCell-Cell_data(:,2);
SquaredResidualsCell = ResidualsCell.*ResidualsCell;

SumSquaredResiduals =
sum(SquaredResidualsSubstrate)+sum(SquaredResidualsCO2)+sum(SquaredResidualsCell)

% Mass balance batch calculation (does not account for leakage)
MolesVwVair = concentrations(:,1)*Capacity(1,1);
MolesBiomass = concentrations(:,2)*Capacity(2,1);
MolesCO2 = concentrations(:,3)*Capacity(3,1);
MolesStart=concentrations(1,:)'.*Capacity;
MolesEnd=concentrations(end,:)'.*Capacity;
MassbalanceBatch = sum(MolesEnd(1:3))/sum(MolesStart(1:3))

figure('Position',[100 198 560 420])
subplot(2,2,1);
plot(time/3600,Cw,Sub_data(:,1)/3600,Sub_data(:,2),'d');
xlabel('t [h]');
ylabel('c [moles/m3]');
title('Substrate C concentration in batch water')

subplot(2,2,2);
plot(time/3600,Cb,Cell_data(:,1)/3600,Cell_data(:,2),'d');
xlabel('t [h]');
ylabel('c [moles/m3]');
title('Biomass C concentration in batch water')

subplot(2,2,3);
plot(time/3600,CO2,CO2_data(:,1)/3600,CO2_data(:,2),'d');
xlabel('t [h]');
ylabel('c [moles/m3]');
title('Batch headspace carbon dioxide C concentration')

subplot(2,2,4);
```

```
plot(time/3600,MolesVwVair,time/3600,MolesBiomass,time/3600,MolesCO2);
xlabel('t [h]');
ylabel('c [moles C]');
legend('Substrate C in water and air','Biomass C','Inorganic C')
title('Moles of C in different compartments')

figure('Position',[500 120 560 420])
subplot(2,2,1);
mesh(umax_range1,KS_range1,SumSquaredResiduals1);
xlabel('umax [s-1]');
ylabel('KS [moles/m3]');
zlabel('Sum Squard Residuals');
title('SumSqResiduals, first iteration')
subplot(2,2,2);
mesh(umax_range2,KS_range2,SumSquaredResiduals2);
xlabel('umax [s-1]');
ylabel('KS [moles/m3]');
zlabel('Sum Squard Residuals');
title('SumSqResiduals, second iteration')
subplot(2,2,3);
contour(umax_range1,KS_range1,SumSquaredResiduals1);
xlabel('umax [s-1]');
ylabel('KS [moles/m3]');
zlabel('Sum Squard Residuals');
title('SumSqResiduals, first iteration')
subplot(2,2,4);
contour(umax_range2,KS_range2,SumSquaredResiduals2);
xlabel('umax [s-1]');
ylabel('KS [moles/m3]');
zlabel('Sum Squard Residuals');
title('SumSqResiduals, second iteration')
```

8. CO_2 Leak Test

Two control experiments were conducted in triplicate to confirm whether or not there was an apparent loss of CO_2 via diffusion through the gap between walls of the vials and the Mininert caps. One batch was tightly capped with mininert caps, whereas the other batch was capped with rubber septum crimped with an aluminium cap. The tightly capped reactor bottles were filled with 200 mL of sterile deionized water and injected with 1.5 mL of pure CO_2 (CP Grade, BOC Gases, Surrey, UK) followed by the monitoring of headspace CO_2 concentrations according to the methods described in **Appendix B (Section 5)** for 6 days.

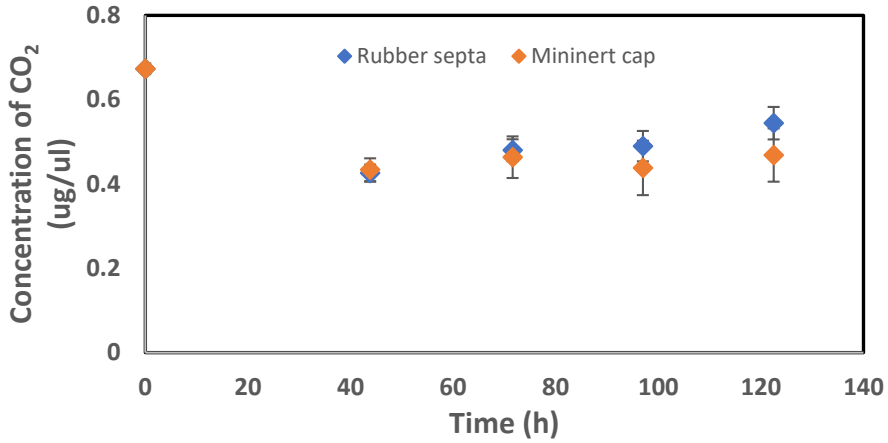


Figure B 1Headspace CO_2 concentration over time for reactor bottles capped with either rubber septa and miniinert cap to evaluate if there is leakage of gas through the bottles over time. Data points are an average of triplicate test for each type of cap used.

9. Stoichiometric Equations and Standard Gibbs free Energy for Test Chemical Biodegradation

The half reaction approach for energy and synthesis reaction

Energy Reaction (R_e)

- Develop a balanced oxidation half reaction (R_d) for the electron donor substrate written on an one-electron equivalent basis
- Write a balanced reduction half reaction for oxygen (electron acceptor), also written for one electron equivalent (R_a)(McCarty, 2012)
- Energy reaction (R_e) = R_a - R_d

Synthesis Reaction (R_s)

- Write a cell half reaction for synthesis (R_c). The selection of this reaction depends on the type of nitrogen source available for the formation of protein and nucleic acids(McCarty, 2012).
- Energy reaction (R_s) = R_c - R_d

The overall reaction [R] which includes both energy generation and synthesis can be written as:

$$R = f_e R_e + f_s R_s$$

Where, f_e and f_s are the fraction of electron equivalents used for energy and synthesis reactions, respectively.

Free energy of the energy reaction

The standard free energies (ΔG_r^0) available from the biodegradation of each test chemical were determined by using the values of free energy of formation ($\mathbf{\Delta G_f^o}$) for individual

constituents, as listed by (McCarty, 2012) using the equation mentioned elsewhere (Dolfing and Janssen, 1994)

a. Phenol

$$0.035714\ C_6H_6O\ +\ 0.15\ O_2 +\ 0.02\ NH_4^+ = 0.02\ C_5H_7O_2N +\ 0.067143\ H_2O +\ 0.114286\ CO_2 + 0.02\ H^+$$

Free energy of the energy reaction for phenol mineralization

Re: $\frac{1}{28} C_6H_6O\ + \frac{1}{4}\ O_2 = \frac{3}{28} H_2O + \frac{6}{28} CO_2$

The free energies of formation for each species are expressed in kJ/mol.

$\frac{1}{28}\ (-72.8) + \frac{1}{4}\ (0) = \frac{3}{28}(-237.178) + \frac{6}{28}$ (-394.359)

$\Delta G_t^o = \Delta G^o_p - \Delta G^o_r$

-3004.876 kJ/mol

b. m-cresol

$$0.029412\ C_7H_8O +\ 0.15\ O_2 +\ 0.02\ NH_4^+ = 0.02\ C_5H_7O_2N +\ 0.077647\ H_2O +\ 0.105857\ CO_2 + 0.02\ H^+$$

Free energy of the energy reaction for m-cresol mineralization

Re: $\frac{1}{34} C_7H_8O + \frac{1}{4}\ O_2 = \frac{2}{17} H_2O + \frac{7}{34} CO_2$

The free energies of formation for each species are expressed in kJ/mol.

$\frac{1}{34}\ (-45.606) + \frac{1}{4}\ (0) = \frac{2}{17}(-237.178) + \frac{7}{34}$ (-394.359)

$\Delta G_t^o = \Delta G^o_p - \Delta G^o_r$

= -3646.84 kJ/mol

c. toluene

$$0.027778\ C_7H_8\ +\ 0.15\ O_2 +\ 0.02\ NH_4^+ = 0.02\ C_5H_7O_2N +\ 0.071111\ H_2O +\ 0.094444\ CO_2 + 0.02\ H^+$$

Free energy of the energy reaction for toluene mineralization

Re: $\frac{1}{36} C_7H_8\ + \frac{1}{4}\ O_2 = \frac{4}{36} H_2O + \frac{7}{36} CO_2$

The free energies of formation for each species are expressed in kJ/mol.

$\frac{1}{36}\ (-127) + \frac{1}{4}\ (0) = \frac{4}{36}(-237.178) + \frac{7}{36}$ (-394.359)

$\Delta G_t^o = \Delta G^o_p - \Delta G^o_r$

= -3582.2268 kJ/mol

d. 2,4-dichlorophenol

$0.041667\ C_6H_4Cl2O + 0.15\ O_2 + 0.02\ NH_4^+ = 0.02\ C_5H_7O_2N + 0.15\ CO2 + 0.08333\ Cl^- + 0.103332\ H^+ + 0.001668\ H_2O$

Free energy of the energy reaction of 2,4-dichlorophenol mineralization

Re: $\frac{1}{24} C_6H_4Cl2O + \frac{1}{4}\ O_2 = \frac{1}{4} CO2 + \frac{1}{24} H_2O + \frac{1}{12} Cl^- + \frac{1}{12}\ H^+$

The free energies of formation for each species are expressed in kJ/mol.

$\frac{1}{24}(-84.4) + \frac{1}{4}(0) = \frac{1}{4}(-394.359) + \frac{1}{24}(-237.178) + \frac{1}{12}(-133.26) + \frac{1}{12}(-39.87)$

$\Delta G_t^o = \Delta G^o_p - \Delta G^o_r$

= -2865.192 kJ/mol

e. 4-chlorophenol

$0.0384\ C_6H_5ClO + 0.15\ O_2 + 0.02\ NH_4^+ = 0.02\ C_5H_7O_2N + 0.13078\ CO2 + 0.0385\ Cl^- + 0.0585\ H^+ + 0.0369\ H_2O$

Free energy of the energy reaction of 4-chlorophenol mineralization

Re: $\frac{1}{26} C_6H_5ClO + \frac{1}{4}\ O_2 = \frac{3}{13} CO2 + \frac{1}{13} H_2O + \frac{1}{26} Cl^- + \frac{1}{26}\ H^+$

The free energies of formation for each species are expressed in kJ/mol.

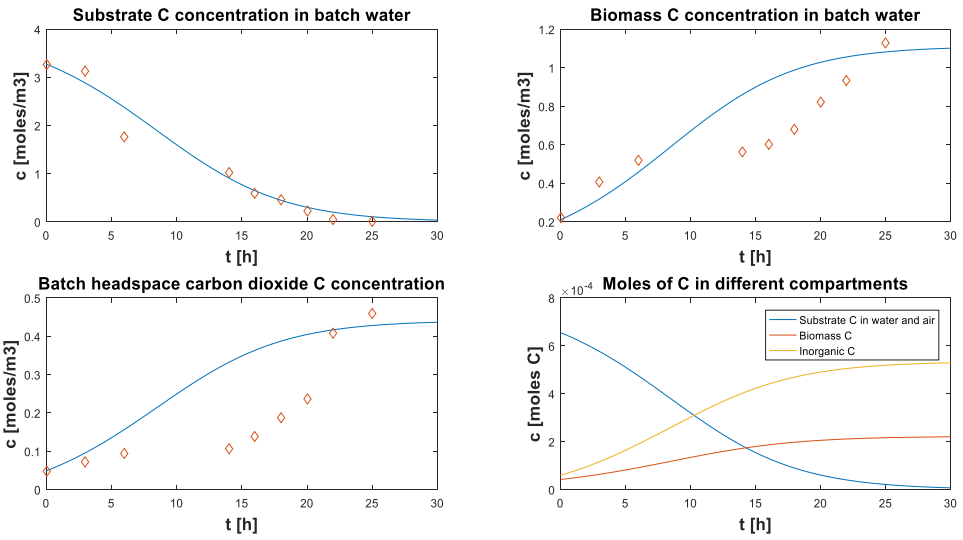
$\frac{1}{26}(-56.8) + \frac{1}{4}(0) = \frac{3}{13}(-394.359) + \frac{1}{13}(-237.178) + \frac{1}{26}(-133.26) + \frac{1}{26}(-39.87)$

$\Delta G_t^o = \Delta G^o_p - \Delta G^o_r$

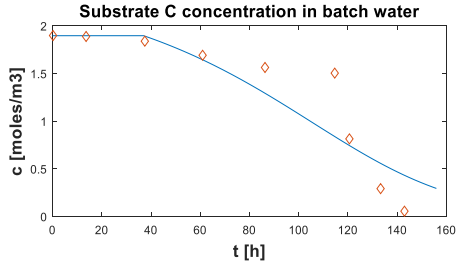
= -2836.496 kJ/mol

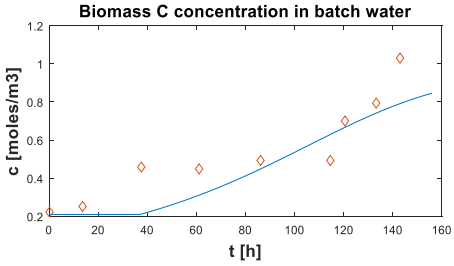
10. Model predicted vs Observed Biodegradation Results

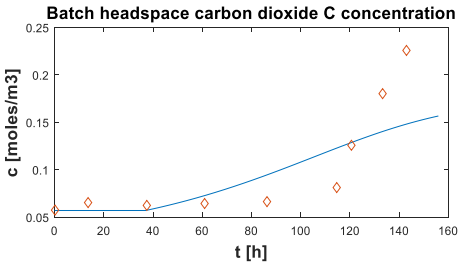
a. m-cresol

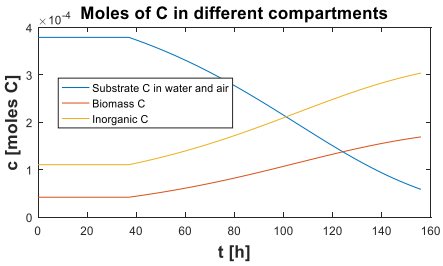


b. 2,4-dichlorophenol

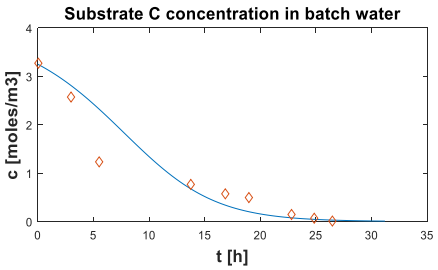


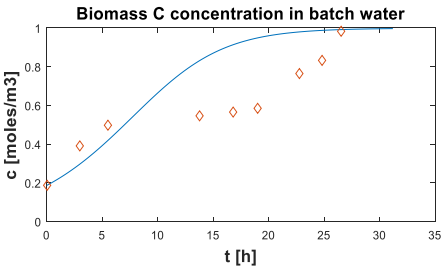


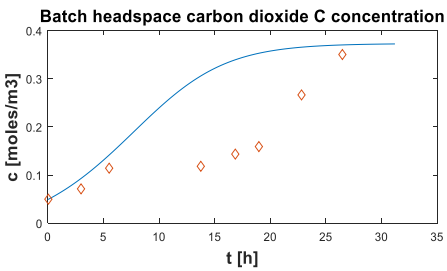


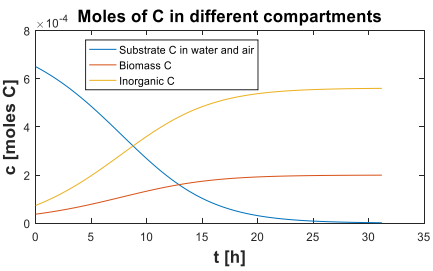


c. Phenol

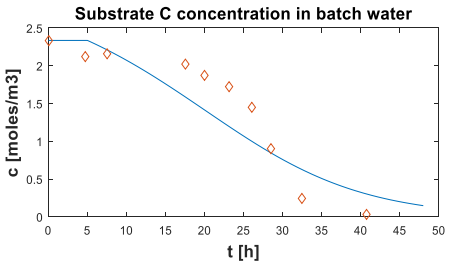


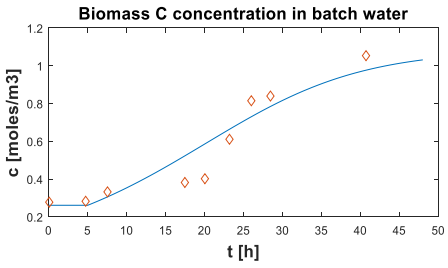


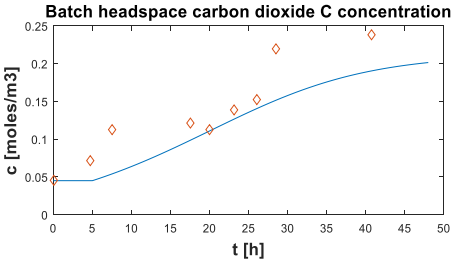


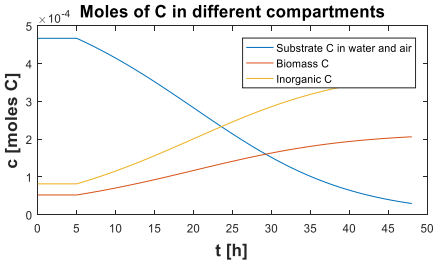


d. 4-chlorophenol









e. Toluene

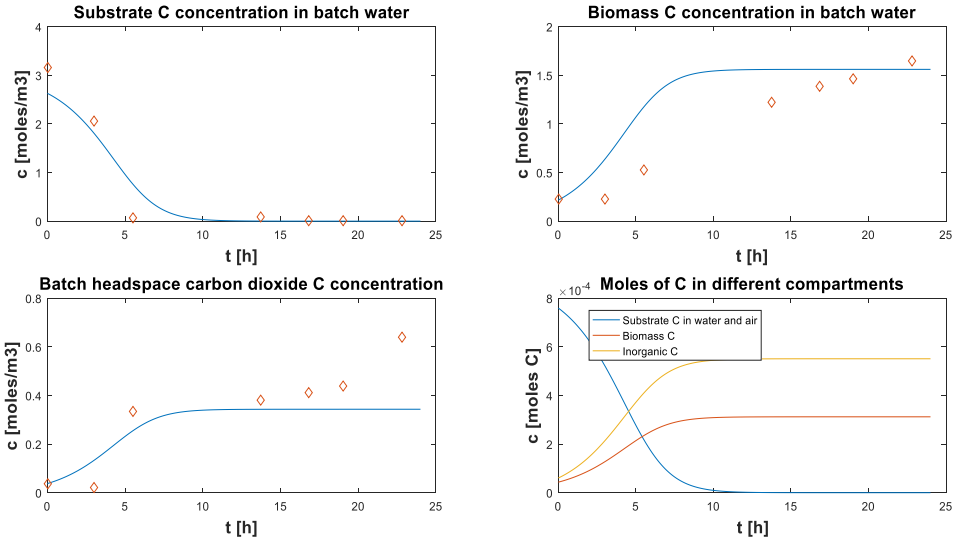


Figure B 2. Monod kinetic model fit to the experimental data on substrate removal, degrader growth and headspace CO_2 from the degradation of test chemicals. Fourth plot (bottom right) from each of the five figures (a,b,c,d &e) is the carbon mass balance that accounts the initial and final carbon present in the biomass, test chemical and inorganic carbon (CO_2) in the batch reactor system.

11. QSBR

Table B 3. Values of selected molecular descriptors and maximum specific substrate removal rate for test chemicals

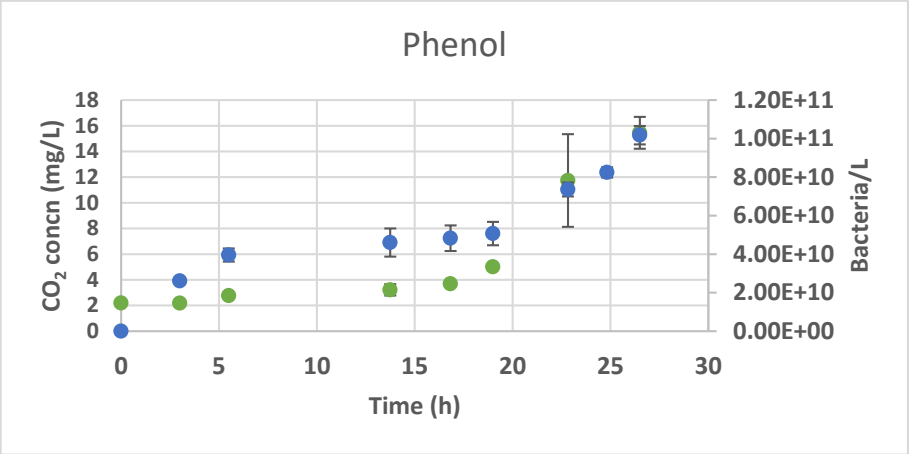
Chemicals	Mor08u	nArX	total energy	E_{homo}	E_{lomo}	Electrical energy	MW	vdw V	MR	Log P	σ	$_vX^1$	$_vX^2$	q_{max}
2,4-DCP	0.126	2	-1619.95	9.58	0.723	-6798.48	163.00	118.4	37.92	2.99	0.54	3.096	2.439	0.090
4-CP	-0.005	1	-1366.57	9.33	0.411	-5498.44	128.55	104.43	33.02	2.43	-0.14	2.612	1.913	0.758
toluene	0.355	0	-967.72	9.22	0.317	-4275	92.14	98.87	31.07	2.68	-0.17	2.411	1.655	2.344
phenol	-0.113	0	-1113.0	9.23	0.036	-4432.63	94.11	90.52	28.13	1.48	-0.37	2.134	1.336	2.743
m-cresol	0.304	0	-1263.19	9.09	0.164	-5254.96	108.13	104.1	32.95	1.94	-0.54	2.545	1.839	2.730

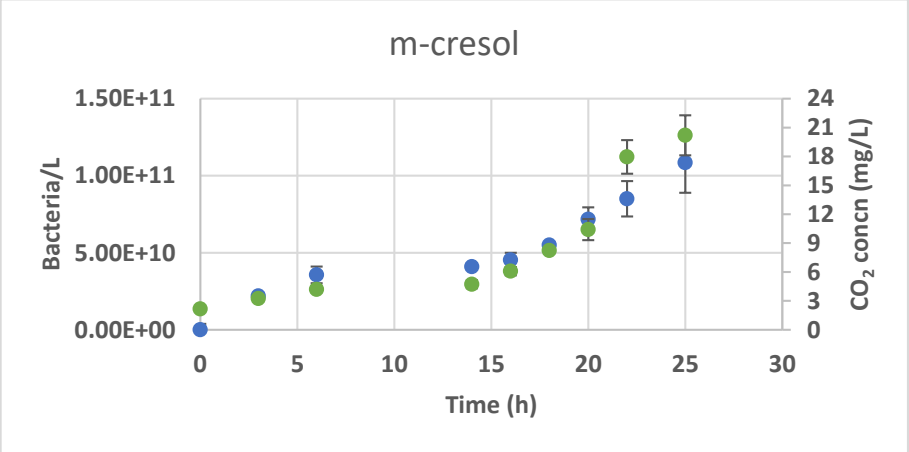
12. Amount Oxygen required for mineralization of test chemicals in the bioreactor

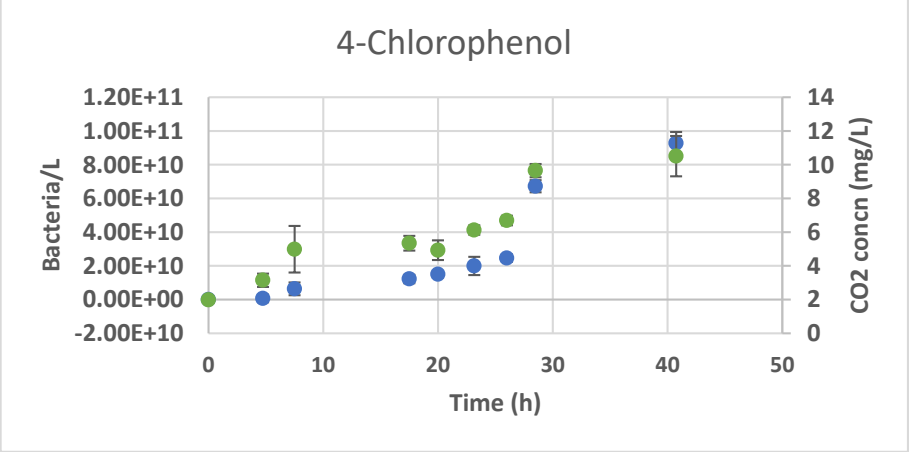
Table B 4. Summary table showing the amount of test chemicals spiked in the bioreactors, oxygen required to aerobically mineralize the spiked test chemicals, available oxygen in the headspace of bioreactor andthe dissolved oxygen content of minimal medium.

Chemicals	Moles of chemical spiked	moles of oxygen required for mineralization	moles of oxygen in HS	DO in DI water (mg/L)
2,4-dichlorophenol	6.13E-05	2.21E-04	3.12E-03	9.68
4-chlorophenol	7.78E-05	3.04E-04	3.12E-03	
toluene	1.09E-04	5.86E-04	3.12E-03	
m-cresol	9.25E-05	4.72E-04	3.12E-03	
phenol	1.06E-04	4.46E-04	3.12E-03	

13. Plot of biomass growth and headspace CO_2 development during the biodegradation of studied chemicals







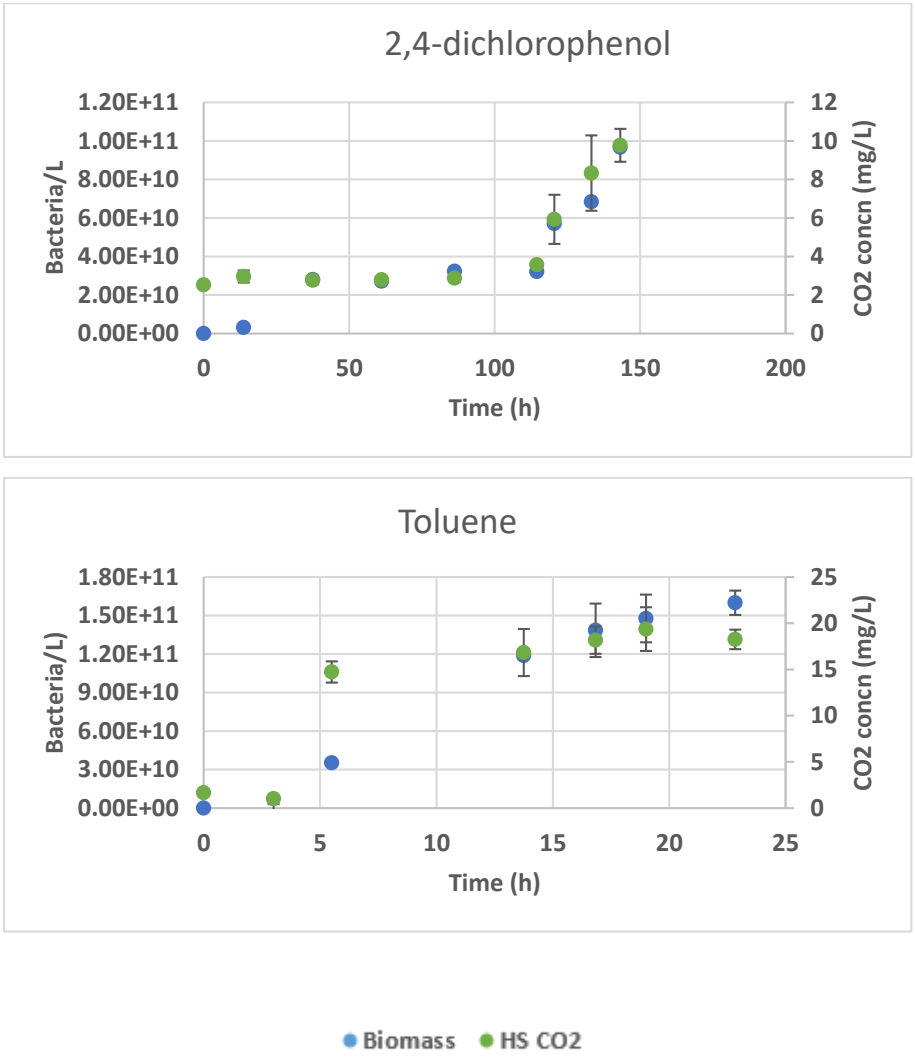


Figure B 3. Biomass growth and headspace CO_2 concentration profile obtained during the biodegradation of test chemicals with known degraders at room temperature. Orange and blue circles respectively, represent the headspace CO_2 concentration for each chemical biodegradation. Error bars represent the standard deviations for triplicate experiments and might not be visible in some cases.

14. Calculation of cell carbon content

The cell carbon content of bacterium during the biodegradation of 2,4-DCP was estimated using the following approach. Similar approach was used to estimate the cell carbon content for other test chemicals biodegradation

Molecular weight of 2,4 DCP = 163

Carbon content (C1) = 72/163 = 0.441

Spiked 2,4-DCP concentration [g/L] (C2) = 0.05

Concentration of carbon [gC/L) (C3) = C1 * C2 = 0.05 * 0.441 = 0.022

Equation for 2,4-DCP mineralization

$$\mathbf{0.041667}\ C_6H_4Cl2O + \mathbf{0.15}\ O_2 + \mathbf{0.02}\ NH_4^+ = \mathbf{0.02}\ C_5H_7O_2N + \mathbf{0.15}\ CO2 + \mathbf{0.08333}\ Cl^- + \mathbf{0.103332}\ H^+ + \mathbf{0.001668}\ H_2O$$

1 mole of 2,4-DCP produced 0.048 moles of biomass and 3.6 moles of CO2

1) Carbon content in biomass $\left(\frac{\text{gC}}{\text{L}}\right) =$

$$\frac{(\text{total no.of carbon in biomass} * \text{number of moles of biomass produced})}{\text{total number of carbon in 2,4-DCP}} * \text{C3}$$

$$\text{Carbon content in biomass } \left(\frac{\text{gC}}{\text{L}}\right) = \ 0.0088$$

2) Carbon content in CO2 $\left(\frac{\text{gC}}{\text{L}}\right)$

$$= \frac{(\text{total no. of carbon in CO2} * \text{number of moles of CO2 produced})}{\text{total number of carbon in } 2,4 - \text{DCP}} * \text{C3}$$

$$\text{Carbon content in CO2 } \left(\frac{\text{gC}}{\text{L}}\right) = \ 0.0132$$

3) Total number of bacteria synthesised $(\frac{\text{bacteria}}{\text{L}})$

$$= \text{Bacteria count (end)} - \textit{Bacteria count (initial)}$$

$$\text{Total number of bacteria synthesised } \left(\frac{\text{bacteria}}{\text{L}}\right) = 9.67 * 10\text{^}10$$

$$\textbf{Bacteria cell carbon content } \left(\frac{\textbf{gC}}{\textbf{cell}}\right) = \frac{\textbf{Carbon content in biomass}}{\textbf{Total number of bacteria synthesised}}$$

$$= \mathbf{9.11681E-14}$$

Appendix C

Appendix C.1

Table C 1 List of sequences used in this study and associated organism. The model organism used for each gene is reported in bold and blue.

Gene	Gene Bank Accession number	Microorganisms
2,4-dichlorophenol hydroxylase (*tfdB*)	**AY365053.1**	***Ralstonia eutropha* JMP134 plasmid pJP4**
	JQ436721.1	*Variovorax sp.* DB1 plasmid pDB1
	JX847411.1	*Burkholderia cepacia* strain 2a plasmid pIJB1
	AB212794.1	*Burkholderia sp*. Ff54
	AB212792.1	*Burkholderia sp*. M701
	AB212785.1	*Ralstonia sp.* T101
	U88192.1	*Burkholderia cepacia* plasmid pIJB1
	AB212790.1	*Burkholderia sp.* T201
	JQ436722.1	
Catechol-2,3-dioxygenase (*C23D*)	**AB910524.1**	***Pseudomonas sp. CF600* plasmid pVI150**
	AY887956.1	*Pseudomonas mendocina* strain PC8
	KU695543.1	*Pseudomonas pseudoalcaligenes* strain C70
	S77084.1	*Alcaligenes*, KF711
	JX177746.1	*Pseudomonas stutzeri* strain 2A1
	X77856.1	*P.putida*
	Y887950.1	*Pseudomonas mendocina* strain PC2
	DQ131592.1	*Pseudomonas putida* strain ATCC 11172
	AB234618.1	*Pseudomonas stutzeri* plasmid NAH
	AY887970.1	*Pseudomonas fluorescens* strain PC36
	AB238971.1	*Pseudomonas putida* plasmid pWW53
	U01825.1	*Pseudomonas sp.* IC
	X60740.1	*Pseudomonas aeruginosa* xylEJI104-1
	M65205.1	DK1XYLE TOL Plasmid pDK1
	V01161.1	*Pseudomonas putida*
Phenol Hydroxylase (*PH*)	JX131378.1	*Rhodococcus sp*. JDD1H
	AB286755.1	Uncultured bacterium
	HQ915639.1	*Pseudomonas sp.* gl06
	DQ852625.1	*Pseudomonas putida* strain PD39
	AY875746.1	*Pseudomonas fluorescens* strain PC36

	AB985593.1	*Marinobacter sp.*
	JQ015321.1	*Burkholderia sp.* HA-OP24
	AB910524.1	***Pseudomonas sp. CF600* plasmid pVI150 DNA**

Appendix C.2

Table C 2 Key catabolic genes and bacterial reference strains used in this study, their growth conditions and origin.

Catabolic genes	Bacteria Strain	Growth Conditions	Reference or Source
2,4-dichlorophenol hydroxylase (*tfdB*), 3,5-dichlorocatechol-1,2-dioxygenase (*tfdC*)	*Ralstonia eutropha JMP134 (DSMZ)*	M9 with 2,4-dichlorophenoxyacetic acid (≥ 98.0% SigmaAldrich), 28^{o}C (Dorn et al 1974)	Benson et al. 2013
Catechol-2,3-dioxygenase (*C23D*), Phenol hydroxylase (*PH*)	*Pseudomonas putida CF600 (Department of Molecular Biology, Umea University, Sweden)*	Luria-Bertani (LB) broth medium, 30 ^{o}C (Sigma Aldrich) (Shingler et al. 1989)	Shingler et al. 1989, Benson et al. 2013

Appendix C.3

Example calculation to estimate gene copies/µL for qPCR standard generation

Phenol hydroxylase (PH primers) amplicon size = 174 bp

Assume average bp weight of 1.079×10^{-21} g/molecule as you do not know the exact sequence composition.

Weight of fragment = bp weight x fragment length (bp)

$= 1.079 \times 10^{-21} \times 174$

$= 1.877 \times 10^{-19}$

$$\text{Finally, Copies/µl} = \frac{\text{Concentration of amplicon (g/µl)}}{\text{Total wt of amplicon (g/molecule)}}$$

The stock solution was preserved in concentration of at least 10^9 copies/µl, and the standard solutions were prepared daily by ten-fold dilutions, in order to create a standard curve in which samples concentration falls within the highest and lowest standard concentration.

Appendix C.4

Table C 3 RNA yield and purity of total RNA isolated from activated sludge inocula over time from phenol and 2,4-DCP biodegradation experiments. The time zero sample for RNA extraction was collected prior spiking the chemicals in the bioreactor.

Sample ID	**RNA concentration after clean up (ng/µL)**	**A $_{260/280}$**	**A $_{260/230}$**
Pseudomonas sp. CF600 in minimal medium with Phenol	395.08	2.15	2.01
Ralstonia eutropha JMP134 in minimal medium with 2,4-DCP	23.4	2.14	1.86
Phenol spiked activated sludge (WWTP1)			
Time (h)	**RNA concentration after clean up (ng/µL)**	**A $_{260/280}$**	**A $_{260/230}$**
0	59.8	2.06	1.89
3	87	2.14	1.83
10	136.1	1.96	1.61
2,4-DCP spiked activated sludge (WWTP4)			
Time (h)	**RNA concentration after clean up (ng/µL)**	**A $_{260/280}$**	**A $_{260/230}$**
0	29.2	2.16	2.03
40	55	2.2	1.74
71	42.1	2.44	1.89
162	53.5	2.25	1.81
2,4-DCP spiked activated sludge (WWTP1)			
Time (h)	**RNA concentration after clean up (ng/µL)**	**A $_{260/280}$**	**A 260/230**
0	29.2	2.16	2.03

Appendix C.5

PCR products sequencing results screen shot

a. Chlorocatechol-1,2-dioxygenase gene (*tfdC*)

Description	Total Score	Query Coverage	E-value	Identity	Accession
Ralstonia eutropha JMP134 plasmid pJP4, complete sequenc	411	100%	0.00E+00	100%	AY365053.1
Uncultured bacterium clone RB3C4 chlorocatechol 1,2-dioxy	383	96%	1.00E-102	98%	AF527740.1
Pandoraea pnomenusa strain MCB032 plasmid unnamed 2, c	377	97%	6.00E-101	97%	CP015373.1
Pandoraea pnomenusa LysR-type transcriptional regulator (	377	97%	6.00E-101	97%	EF600715.1
Uncultured bacterium clone RB3C3B chlorocatechol 1,2-diox	377	96%	6.00E-101	97%	AF527745.1
Uncultured bacterium clone RB3C7 chlorocatechol 1,2-dioxy	377	96%	6.00E-101	97%	AF527743.1
Uncultured bacterium clone RB3C1 chlorocatechol 1,2-dioxy	377	96%	6.00E-101	97%	AF527737.1
Pseudomonas aeruginosa chlorocatechol-1,2-dioxygenase (	377	97%	6.00E-101	97%	AF161263.1
Paraburkholderia xenovorans LB400 chromosome 1, comple	372	97%	3.00E-99	96%	CP008760.1
Pseudomonas knackmussii B13 complete genome	372	97%	3.00E-99	96%	HG322950.1
Pseudomonas putida clc transposon	372	97%	3.00E-99	96%	AJ617740.2
Bordetella petrii strain DSM 12804, complete genome	372	97%	3.00E-99	96%	AM902716.1
Alcaligenes sp. NyZ215 LysR-type regulatory protein (clcR), c	372	97%	3.00E-99	96%	EF544605.1
Paraburkholderia xenovorans LB400 chromosome 1, comple	372	97%	3.00E-99	96%	CP000270.1
Pseudomonas aeruginosa chlorocatechol-1,2-dioxygenase (	372	97%	3.00E-99	96%	AF164958.1
Plasmid pAC27 (from Escherichia coli) catechol oxygenase II	372	97%	3.00E-99	96%	M16964.1
Pseudomonas aeruginosa Lys-R type regulatory protein (clcF	368	97%	3.00E-98	96%	AF087482.1
Ralstonia sp. JS705 clcR, clcA, mcbF, mcbAa, mcbAb, mcbAc,	366	97%	1.00E-97	96%	AJ006307.2
Plasmid pAC27 (from Pseudomonas putida) pyrocatechase (	366	97%	1.00E-97	96%	M36279.1
Uncultured bacterium clone RB3C5 chlorocatechol 1,2-dioxy	355	90%	3.00E-94	98%	AF527741.1
Burkholderia cepacia strain WZ1 chlorocatechol 1,2-dioxyge	351	95%	3.00E-93	95%	EU586138.1
Acidovorax sp. JS773 chlorocatechol 1,2-dioxygenase gene, p	346	89%	2.00E-91	97%	DQ146633.1
Ralstonia sp. T101 tfdC2 gene for chlorocatechol 1,2-dioxyge	346	89%	2.00E-91	97%	AB212978.1
Ralstonia sp. JS704 chlorocatechol 1,2-dioxygenase gene, pa	340	89%	8.00E-90	97%	DQ146632.1
Uncultured bacterium clone CCJA1 chlorocatechol 1,2-dioxy	339	96%	3.00E-89	94%	GU936722.1
Ralstonia sp. C-0 chlorocatechol 1,2-dioxygenase (clcA) gene	335	87%	4.00E-88	97%	AF077914.1
Uncultured bacterium clone CCJB3 chlorocatechol 1,2-dioxy	333	96%	1.00E-87	94%	GU936729.1

b. 2,4-dichlorophenol hydroxylase gene (*tfdB*)

Description	Total Score	Query Coverage	E-value	Identity	Accession
Ralstonia eutropha JMP134 plasmid pJP4, complete sequence	1537	100%	0	100%	AY365053
Ralstonia eutropha LysR-type transcriptional regulator (tfdT), chloropheno	1537	100%	0	100%	U16782.1
Uncultured bacterium plasmid pEMT3, complete sequence	1531	100%	0	99%	JX469827.
Ralstonia pickettii plasmid p712, complete sequence	1531	100%	0	99%	JQ436722.
Variovorax sp. DB1 plasmid pDB1, complete sequence	1110	100%	0	91%	JQ436721.
Burkholderia cepacia plasmid pIJB1 2,4-dichlorophenol hydroxylase (IJBB)	861	100%	0	85%	U88192.1
Uncultured bacterium clone ral 2,4-dichlorophenol hydroxylase (tfdB) gen	846	57%	0	99%	EU429461
Uncultured bacterium partial tfdB gene for 2,4-dichlorophenol hydroxylas	737	48%	0	99%	LN558730

c. Phenol hydroxylase gene (PH)

B	C	D	E	F	G
Description	**Total Score**	**Query Coverage**	**E-value**	**Identity**	**Accession**
Uncultured bacterium partial LmPH gene for phenol hydroxylase large subunit, clone W	588	98%	0.00	99%	LT604167.1
Uncultured bacterium partial LmPH gene for phenol hydroxylase large subunit, clone W	588	98%	0.00	99%	LT604151.1
Pseudomonas pseudoalcaligenes strain C70 phe operon, partial sequence	588	98%	0.00	99%	KU695543.1
Pseudomonas sp. CF600 plasmid pVI150 DNA, dmpRKLMNOPQBCDEFGHI gene cluster,	588	98%	0.00	99%	AB910524.1
Pseudomonas pseudoalcaligenes strain XY4_phe1 multicomponent phenol hydroxylase	588	98%	0.00	99%	KJ174598.1
Bacterium enrichment culture clone PT1 phenol hydroxylase gene, partial cds	588	98%	0.00	99%	KJ004664.1
Pseudomonas pseudoalcaligenes strain C70 multicomponent phenol hydroxylase large s	588	98%	0.00	99%	JX177804.1
Pseudomonas sp. CO-44 phenol hydroxylase gene, partial cds	588	98%	0.00	99%	HM231138.1
Pseudomonas sp. CO-1 phenol hydroxylase gene, partial cds	588	98%	0.00	99%	HM231137.1
Pseudomonas putida strain JFL2008 phenol hydroxylase gene, partial cds	588	98%	0.00	99%	GU047339.1
Pseudomonas putida strain PD39 phenol hydroxylase large subunit gene, partial cds	588	98%	0.00	99%	DQ852625.1
Pseudomonas putida phenol hydroxylase gene cluster (A1, A2, A3, A4, A5, A6), complet	588	98%	0.00	99%	D28864.1
Pseudomonas putida phenol hydroxylase (dmpKLMNOP) operon, complete cds	588	98%	0.00	99%	M60276.1
Uncultured bacterium partial LmPH gene for phenol hydroxylase large subunit, clone W	582	98%	0.00	99%	LT604171.1
Pseudomonas aeruginosa LmPHAT2 gene for phenol hydroxylase largest subunit, partia	582	98%	0.00	99%	AB474002.1

d. Catechol-2,3-dioxygenase gene (C23D)

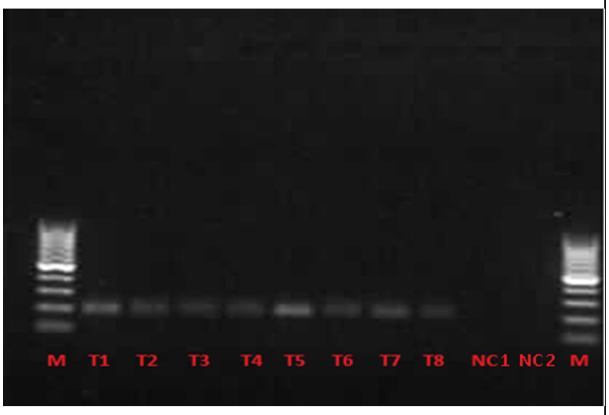
B	C	D	E	F	G
Description	**Total Score**	**Query Coverage**	**E-value**	**Identity**	**Accession**
Pseudomonas putida DNA, complete genome, strain: KF715	267	93%	9.00E-68	96%	AP015029.1
Pseudomonas plecoglossicida strain NyZ12, complete genome	267	93%	9.00E-68	96%	CP010359.1
Pseudomonas sp. CF600 plasmid pVI150 DNA, dmpRKLMNOPQBCD	267	93%	9.00E-68	96%	AB910524.1
Pseudomonas stutzeri strain C52c catechol 2,3-dioxygenase gene, pa	267	93%	9.00E-68	96%	JX177739.1
Planococcus sp. S5 catechol 2,3-dioxygenase gene, complete cds	267	93%	9.00E-68	96%	HQ223337.2
Pseudomonas sp. PD2 catechol 2,3-dioxygenase gene, complete cds	267	93%	9.00E-68	96%	EF635413.1
Pseudomonas putida catechol 2,3-dioxygenase gene, complete cds	267	93%	9.00E-68	96%	EF607052.1
Pseudomonas sp. PD10 catechol 2,3-dioxygenase gene, complete cd	267	93%	9.00E-68	96%	EF607051.1
Pseudomonas sp. PD7 catechol 2,3-dioxygenase gene, complete cds	267	93%	9.00E-68	96%	EF607050.1
Pseudomonas aeruginosa catechol 2,3-dioxygenase (pheB) gene, cor	267	93%	9.00E-68	96%	AY112717.1
Pseudomonas putida strain ATCC 11172 plasmid catechol-2,3-dioxyg	267	93%	9.00E-68	96%	DQ131592.1
Pseudomonas sp. 3YdBTEX2 partial C23O gene for catechol 2,3 diox	267	93%	9.00E-68	96%	AJ544934.1

Figure C 1 The screen shot of closest matching sequence identity of sequenced PCR products. The sequences obtained from sequencing analysis were compared against the Genebank database using BLAST algorithm.

Appendix C.6

Gel electrophoresis results to verify primer specificity

1. **Agarose gel electrophoresis of PH gene: annealing temperature optimization of PH primer set.**



M : DNA marker

T1: PhH gene from Pseudomonas sp. CF600 65 °C

T2: PhH gene from Pseudomonas sp. CF600 64.6 °C

T3: PhH gene from Pseudomonas sp. CF600 63.9 °C

T4: PhH gene from Pseudomonas sp. CF600 62.5 °C

T5: PhH gene from Pseudomonas sp. CF600 60.8°C

T6: PhH gene from Pseudomonas sp. CF600 59.4°C

T7: PhH gene from Pseudomonas sp. CF600 58.5 °C

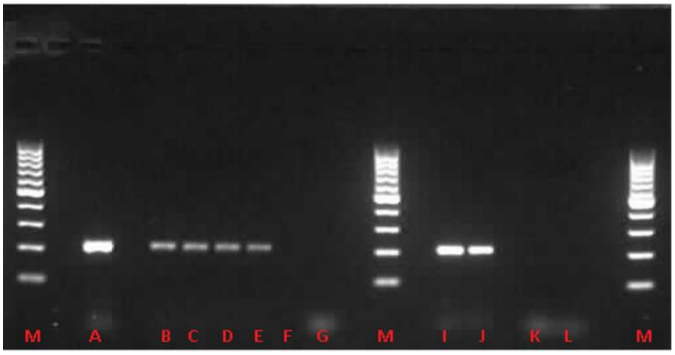
T8: PhH gene from Pseudomonas sp. CF600 58 °C

NC1: Negative control (with water)

NC2 : Negative control (non-target control)

M: DNA ladder

2. **Agarose gel electrophoresis of *C23D* gene and transcript amplified from gDNA isolated from activated sludge and the reference microorganism, respectively. Lane F (water), G (non-target control), K (DNA contamination) and L (water) are different controls.**



M : DNA marker

A: C23D gene in *Pseudomonas sp. CF600*

B,C,D & E: C23D gene in DNA extracted from activated sludge samples

F: Negative control (with water)

G: Negative control for cross- reaction checking

I & J: Positive control (cDNA of C23D transcript from *Pseudomonas sp. CF600*)

K: Control for DNA contamination

L: Negative control (with water)

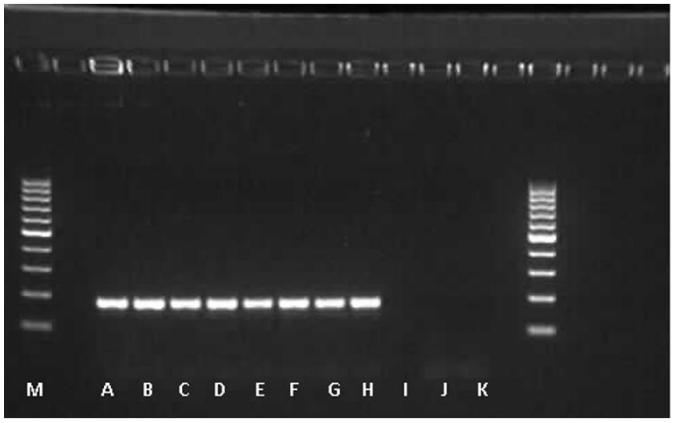
3. **Agarose gel electrophoresis of *PH* transcript amplified from the reference microorganism.**

M : DNA marker

A: Control for DNA contamination

B & C: cDNA of PhH transcripts from *Pseudomonas sp. CF600*

4. **Agarose gel electrophoresis of *PH* gene amplified from total DNA isolated from activated sludge of different WWTPs and the reference microorganism. Lane I (water) and J & K (non-target controls) different controls.**



M : DNA marker

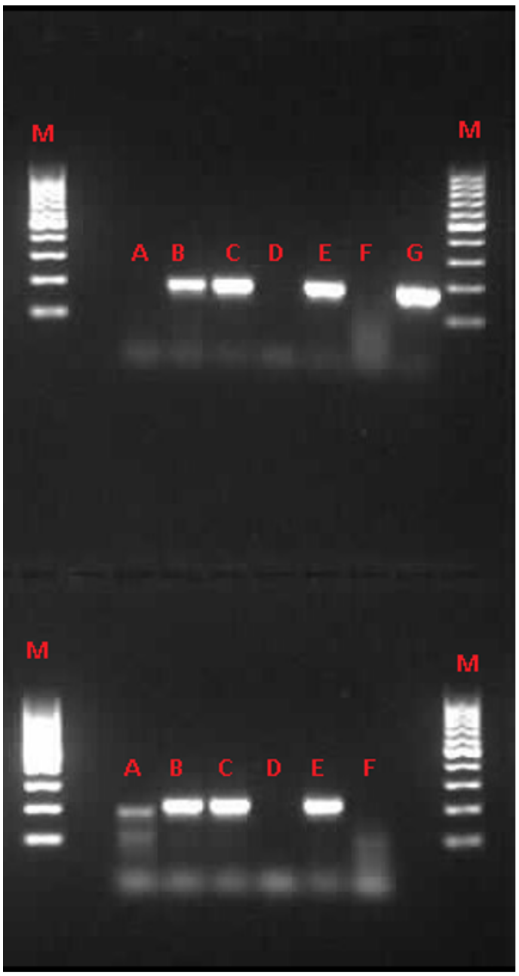
A, B, C, D: PhH gene from *Pseudomonas sp. CF600*

E, F,G & H : PhH gene in DNA extracted from WWTP1, 2, 3 & 4 respectively

I: Negative control (with water)

J & K: Negative control for cross reaction checking

5. Agarose gel electrophoresis : *PH* and *C23D* gene expression during phenol biodegradation studies in activated sludge from WWTP1



M : DNA marker

A: cDNA (PhH gene) at t=0 h

B: cDNA (PhH gene) at t=3 h

C: cDNA (PhH gene) at t=10 h

D: Negative control

E: Positive control (cDNA of PhH transcript from *Pseudomonas sp. CF600*)

F: Control for DNA contamination

G: Standard (10^7 cDNA PhH gene copies/µL)

M : DNA marker

A: cDNA (C23D gene) at t=0 h

B: cDNA (C23D gene) at t=3 h

C: cDNA (C23D gene) at t=7.5 h

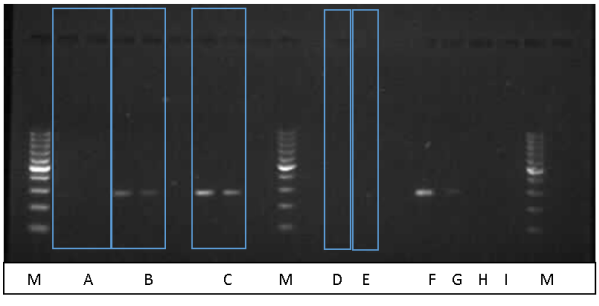
D: Negative control

E: Positive control (cDNA of C23D transcript from *Pseudomonas sp. CF600*)

F: Control for DNA contamination

G: Standard (10^7 cDNA C23D gene copies/µL)

6. Agarose gel electrophoresis: *tfdC* gene in total total DNA extracted from activated sludge of different WWTPs. Lane I (water) and H (non-target control are different negative controls.



M : DNA marker

A: WWTP3

B: WWTP1 + Spiked degrader (*Ralstonia eutropha. JMP134*)

C: WWTP4

D: WWTP1

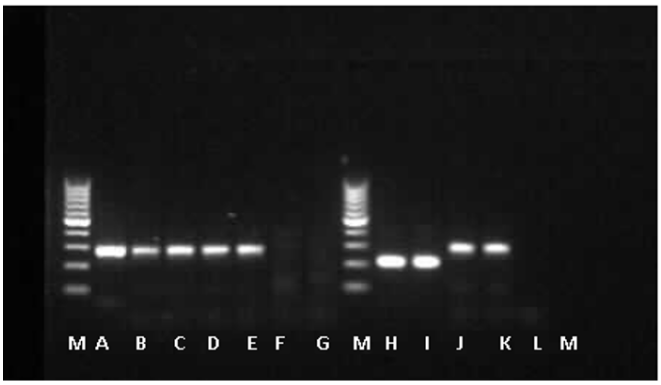
E: WWTP2

F & G: Positive control (*Ralstonia eutropha. JMP134*)

H: Negative control (for cross-reaction checking)

I: Negative control (with water)

7. **Agarose gel electrophoresis: *tfdB* and *tfdC* genes and expression of *tfdC* gene during degradation of 2,4-DCP in activated sludge from WWTP4. Lane L (non-target control for *tfdB*) and M (water) are different negative controls.**



M : DNA marker

A: Positive control (cDNA of tfdC transcript from Ralstonia sp. JMP 134)

B: cDNA (tfdC gene) at t=0 h

C: cDNA (tfdC gene) at t=40 h

D: cDNA (tfdC gene) at t=76 h

E: cDNA (tfdC gene) at t=162 h

F: Negative control (with water)

G : Control for DNA contamination

H: tfdB gene in Ralstonia sp. JMP 134

I: tfdB gene in Ralstonia sp. JMP 134

J: tfdC gene in Ralstonia sp. JMP 134

K: tfdC gene in Ralstonia sp. JMP 134

L: Negative control (tfdB cross- reaction checking)

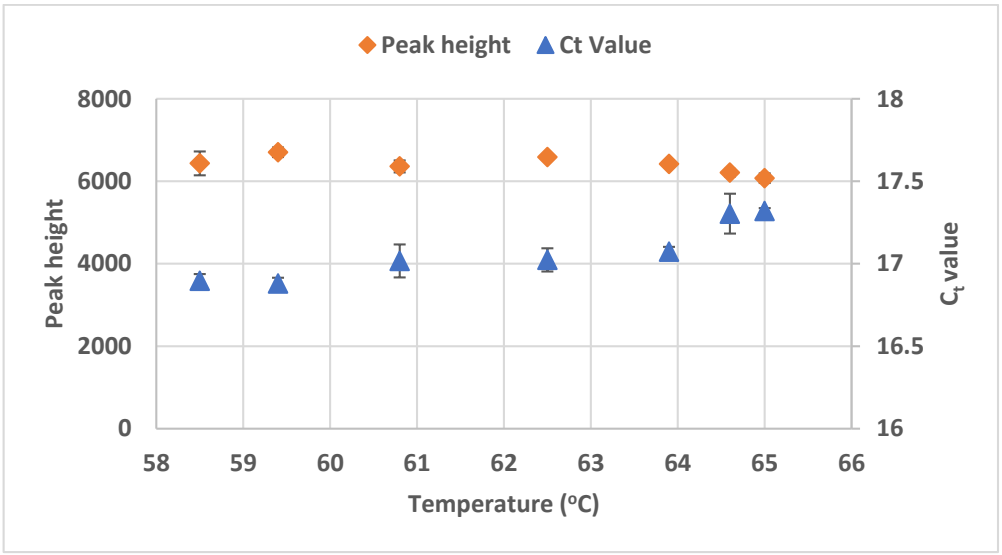
M: Negative control (with water)

Figure C 2 Cross-reactions and specificity of the primer-sets evaluated by agarose gel electrophoresis. Markers correspond to 100-1000-bp PCR marker (Invitrogen, Life Technologies, Paisley, UK).

Appendix C.7

Annealing temperature optimization for *PH* gene (qPCR)

a)



b)

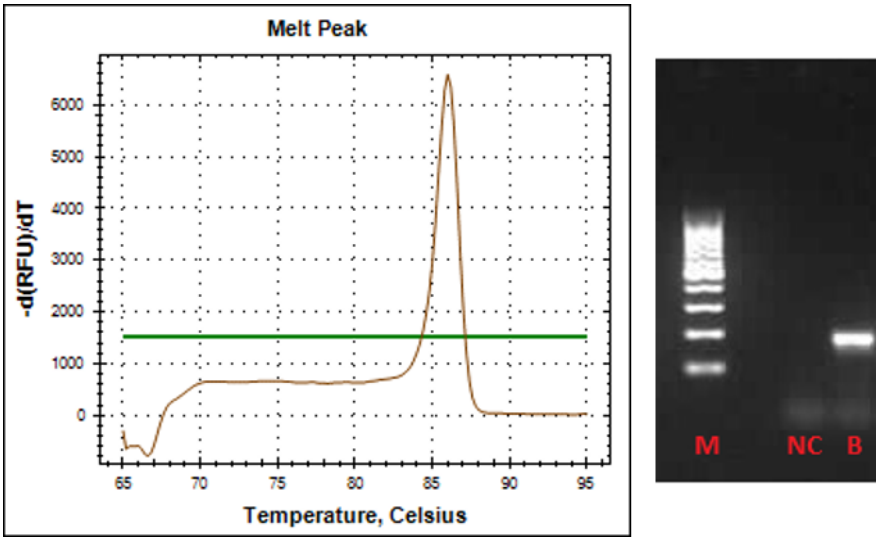
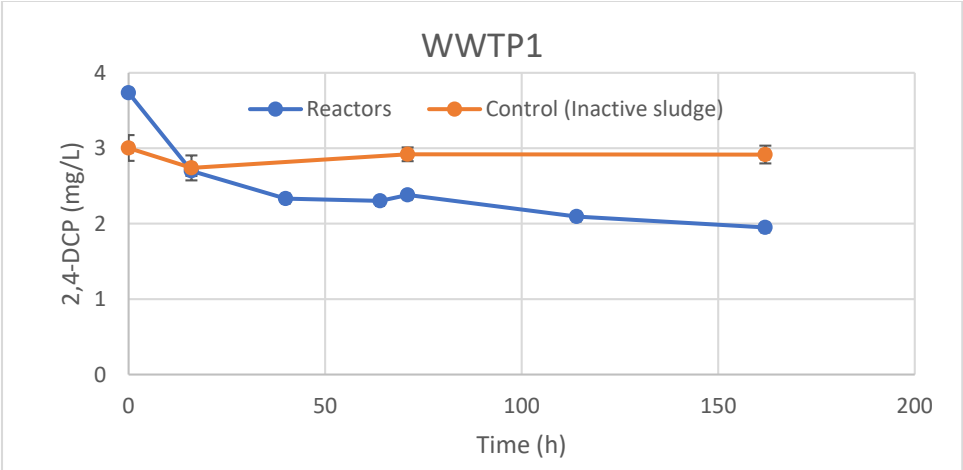


Figure C 3 Optimization of annealing temperature of PH primer set for real-time PCR condition. a) Annealing temperature study: 59.4 ºC reaction gave the lowest C_t value and the highest peak in the melt curve analysis for this assay; b) Melt curve analysis: plot of negative first derivative of the change in fluorescence as a function of temperature, where the peak indicates the melting temperature Tm of a specific PCR product. On the right, the corresponding agarose gel analysis of the real time PCR product: lane M, 100 – 1000 bp PCR marker (Invitrogen, Life Technologies, Paisley, UK); lane NC, negative control with water and lane B, real-time PCR product with PH primer set from the reaction annealing temperature T=59.4 ºC.

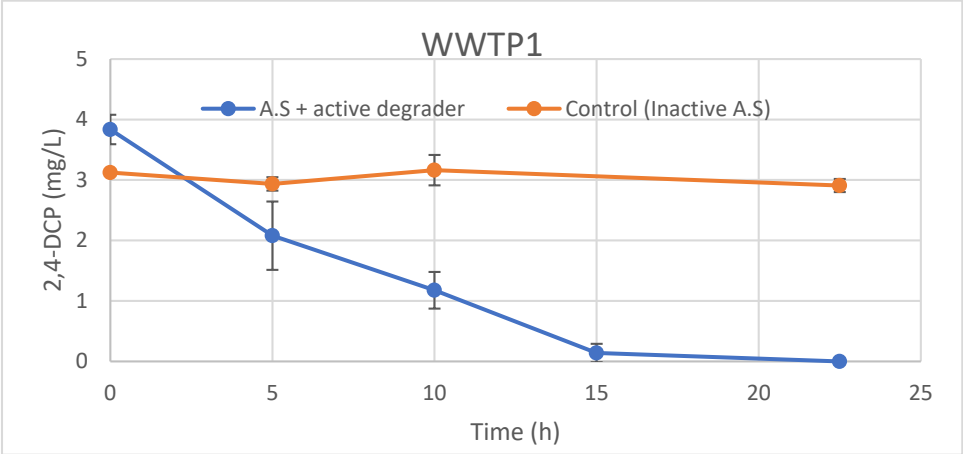
Appendix C. 8

Preliminary 2,4-DCP Biodegradation result

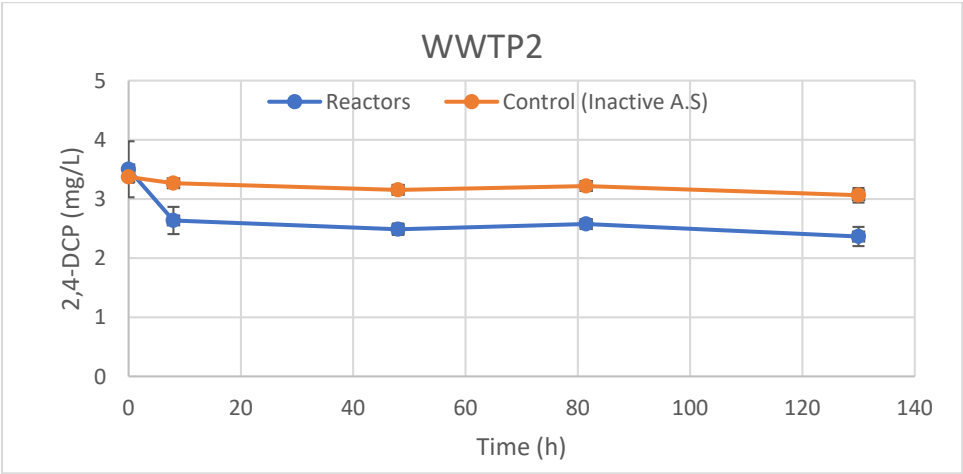
a)



b)



c)



d)

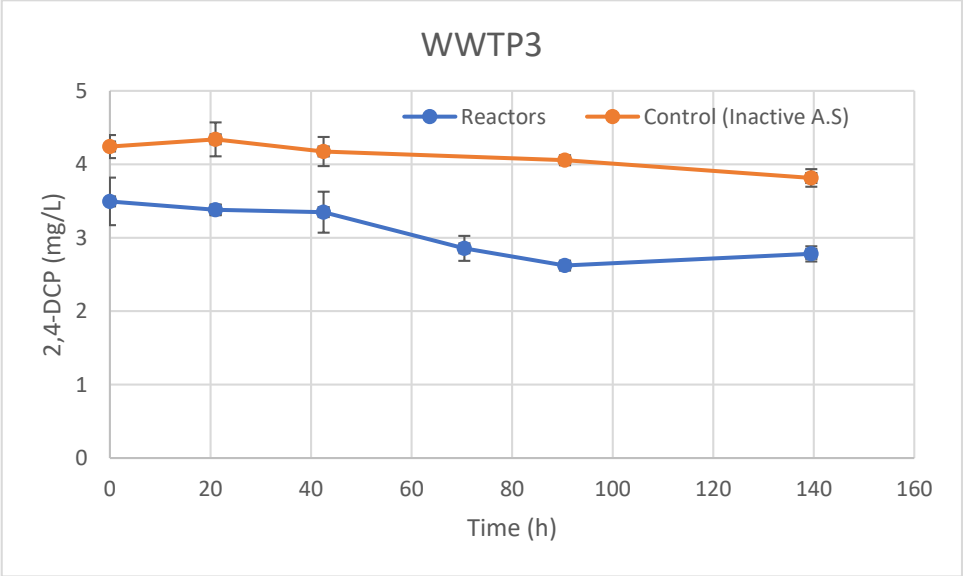
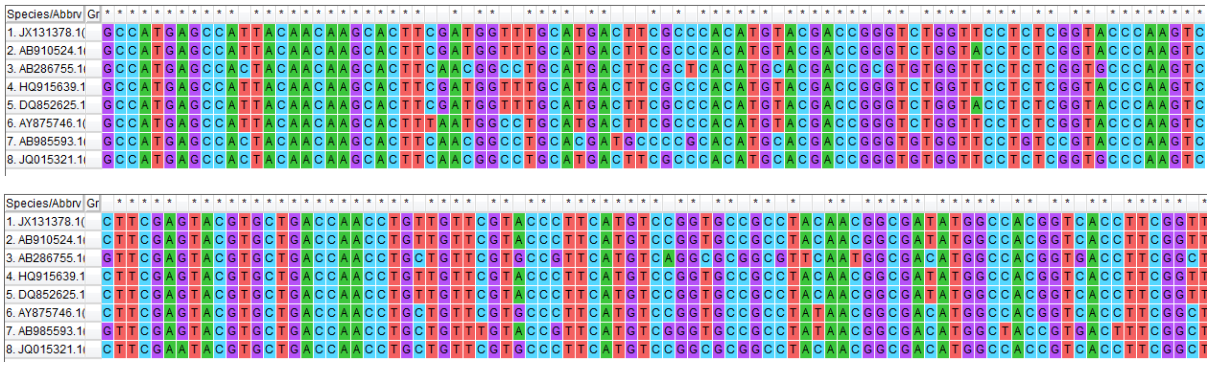


Figure C 4 Trial biodegradation test for 2,4-DCP in activated sludge sampled from different WWTPs (a,b,c,d).Experimentswere conducted in duplicate along with control to check adsorption of 2,4-DCP in solids. For control experiment, single reactor was run, while sampling was perfored in duplicate at each time point. Error bars represent the standard deviations for experiments and might not be visible in some cases

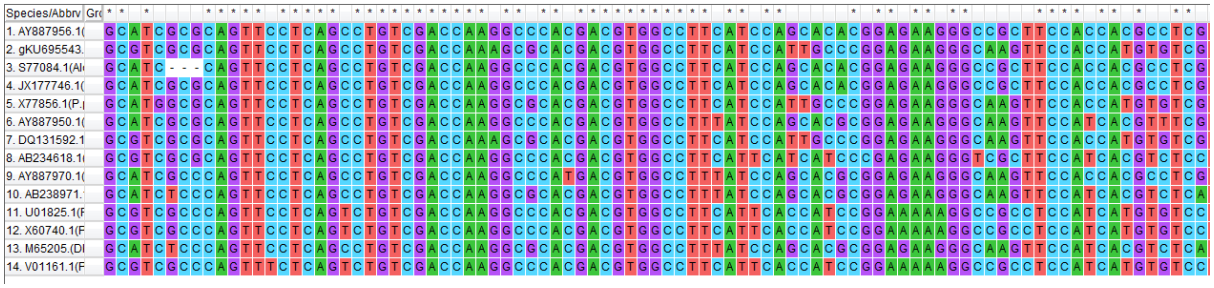
Appendix C.9

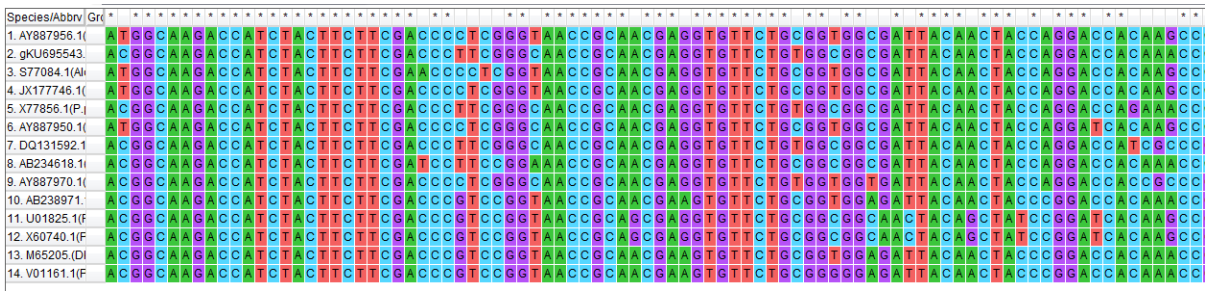
Primer design and gene alignment

1. Phenol Hydroxylase gene alignment

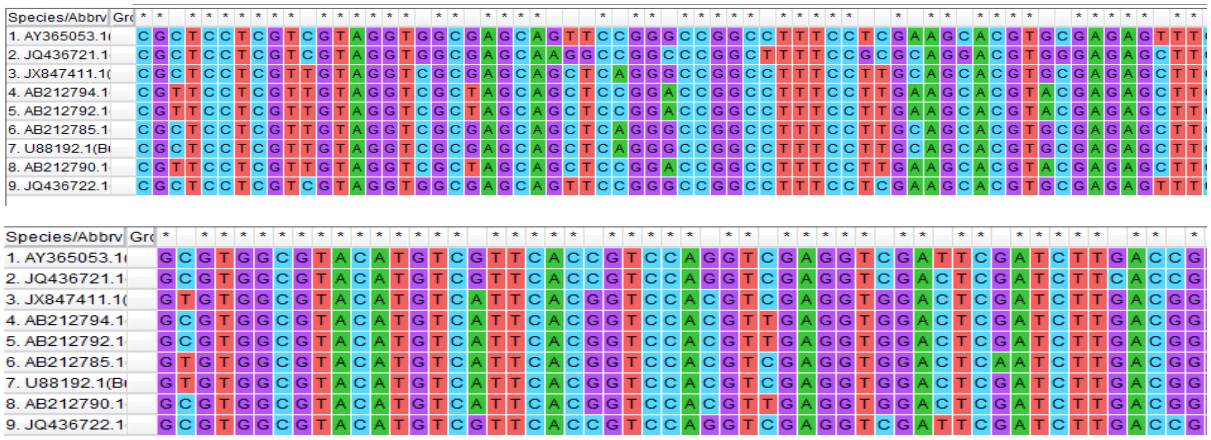


2. Catechol-2,3-dioxygenase gene alignment





3. 2,4-dichlorophenol hydroxylase gene alignment



Appendix C.10

DNA and RNA Extraction

The basic steps in DNA or RNA isolation are;

(1) Disruption (mechanical or chemical) of the cellular structure to create a lysate,

(2) Separation of the soluble DNA or RNA from cell debris and other insoluble material

(3) Purification of the DNA or RNA of interest from soluble proteins and other nucleic acids.

Disruption of most cells in a sample is done by mechanical means or with chaotropic salts, detergents or alkaline denaturation. The lysate is cleared by centrifugation, filtration or magnetic clearing. DNA or RNA soluble portion of the lysate is purified in silica matrices with different washing solutions and finally the DNA or RNA is eluted in an aqueous buffer such as TE buffer or nuclease-free water.

The majority of commercial DNA or RNA isolation systems are based on purification by silica. The DNA or RNA of interest can be extracted by virtue of their ability to bind silica matrix in the presence of high concentrations of chaotropic salts. These salts are then removed with an alcohol-based wash and the DNA or RNA eluted in a low-ionic-strength solution such as TE buffer or water. The binding of DNA or RNA to silica seems to be driven by dehydration and hydrogen bond formation, which competes against weak electrostatic

repulsion. Hence, a high concentration of salt will promote DNA or RNA adsorption onto silica, and a low concentration will release the DNA or RNA.

PureYield™ Plasmid Miniprep System (Promega, UK)

The separation of plasmid DNA from the chromosomal DNA and cellular RNA of host bacteria is the primary consideration of plasmid isolation kit. This system uses SDS alkaline denaturation method, which generates a cleared lysate solution (containing free plasmid) devoid of protein, lipids and chromosomal DNA. This method exploits the difference in denaturation and renaturation characteristics of covalently closed circular plasmid DNA and chromosomal DNA fragments. At alkaline conditions (i.e. pH 11), the plasmid and chromosomal DNA will denaturate. Subsequent, rapid neutralization with a high-salt buffer (e.g., potassium acetate) in the presence of SDS will produce two effects that contribute to the overall effectiveness of the method. First, rapid neutralization causes the chromosomal DNA to base-pair in an intrastrand manner, and ultimately form an insoluble aggregate that precipitates out of solution. Whereas, due to covalently closed nature of the circular plasmid DNA, interstrand rehybridization is achieved, allowing the plasmid to remain in solution. Second, the potassium salt of SDS assists the protein and detergent precipitation and aggregation that promotes the entrapment of the high-molecular-weight chromosomal DNA. Finally the separation of soluble (plasmid DNA) and insoluble material is performed in silica matrix and washing steps is performed remove salts and any other impurities. High quality Plasmid DNA bound in silica column is finally eluted in RNase free water.

ChargeSwitch® gDNA Mini Bacteria Kit (Thermo Fisher Scientific, UK)

This kit is used to isolate high quality genomic DNA from bacterial cells**.** The samples are chemically lysed and homogenized with lysozyme. After preparing the lysates, the gDNA can be purified using this kit, which is based on ChargeSwitch® Technology. This is a novel magnetic bead-based technology that provides a switchable surface which is charge dependent on the pH of the surrounding buffer to facilitate nucleic acid purification. At low pH conditions, there is a positive charge on Magnetic Beads and as a result, negatively charged nucleic acid backbone binds with beads. Proteins and other contaminants are not bound and are simply washed away in an aqueous wash buffer. To elute nucleic acids, the charge on the surface of the bead is neutralized by raising the pH to 8.5 using a low salt elution buffer. Purified DNA elutes instantly into this elution buffer, and is ready for use in downstream applications.

Fast Soil DNA extraction kit (MP Biomedicals, UK)

This kit is used to isolate genomic DNA from wide variety of sources such as soil and other environmental samples (e.g. , activated sludge). The samples are mechanically lysed and homogenized with the FastPrep® Instruments, which uses a unique, optimized motion to homogenize samples by multidirectional, simultaneous impaction with lysing matrix particles. Samples are placed into 2.0 ml tubes containing Lysing Matrix E, a mixture of ceramic and silica particles designed to efficiently lyse all activated sludge organisms including historically difficult sources such as eubacterial spores and endospores, gram positive bacteria, yeast, algae, nematodes and fungi. Homogenization in the FastPrep® Instrument with Lysing Matrix E takes place in the presence of MT Buffer and Sodium Phosphate Buffer, reagents carefully developed to protect and solubilize nucleic acids and proteins upon cell lysis. After lysing step, samples are centrifuged to pellet activated sludge organic maters, cell derbies, and lysing matrix. Proteins from the samples are precipitated with protein precipitating solution (Protein Precipitation Solution is a high-salt buffer that lowers the solubility of proteins). DNA is purified from the supernatant with a silica-based GENECLEAN® procedure using SPIN filters. Any other impurities such as salts, metabolites and proteins are effectively removed by washing using different washing buffers. DNA bound in silica column is finally eluted in RNase free water.

RNA extraction from ISOLATE II Mini RNA kit (Bioline, UK)

This kit is used to isolate high quality total cellular RNA from wide variety of sources including cells and tissues from bacteria, yeast and biological liquid samples. The samples are chemically lysed and homogenized in the presence of guanidinium thiocyanate, a chaotropic salt which immediately deactivates endogenous RNases to ensure purification of intact total RNA. This is followed by addition of ethanol and the sample is than processed through a spin column containing silica membrane, to which RNA binds. The gDNA contamination is remove by an on-column DNase I digestion, any other impurities such as salts, metabolites and cellular components are effectively removed by washing using different buffers. High quality RNA bound in silica column is finally eluted in RNase free water.

Appendix C.11

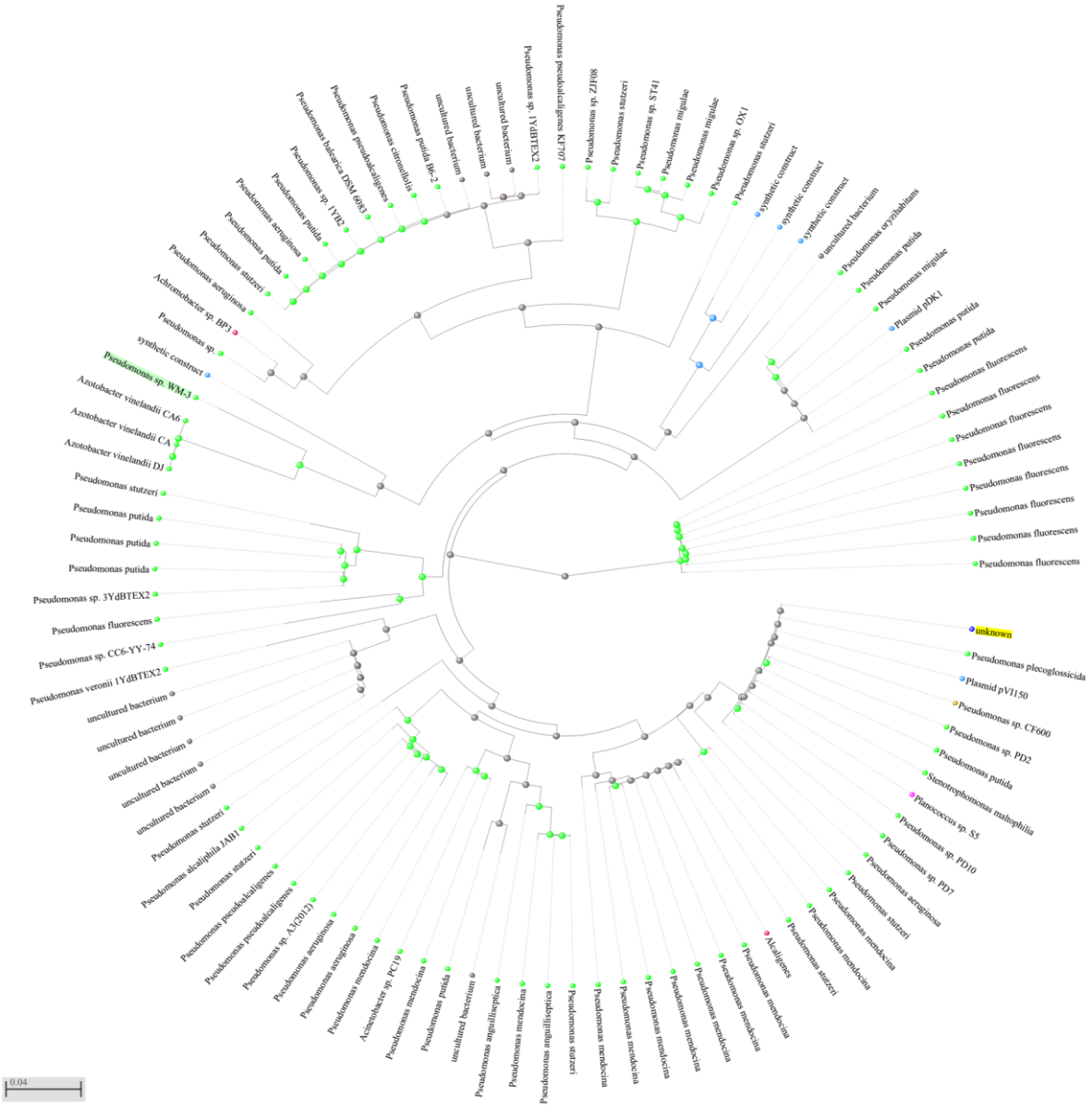


Figure C 5 Neighbour joining phylogenetic tree showing the list of microorganisms that the primer set designed for catechol-2,3-dioxygenase (C23D) targets. These list are obtained from GeneBank database, using the BLAST (Basic Local Alignment Search Tool) algorithm.

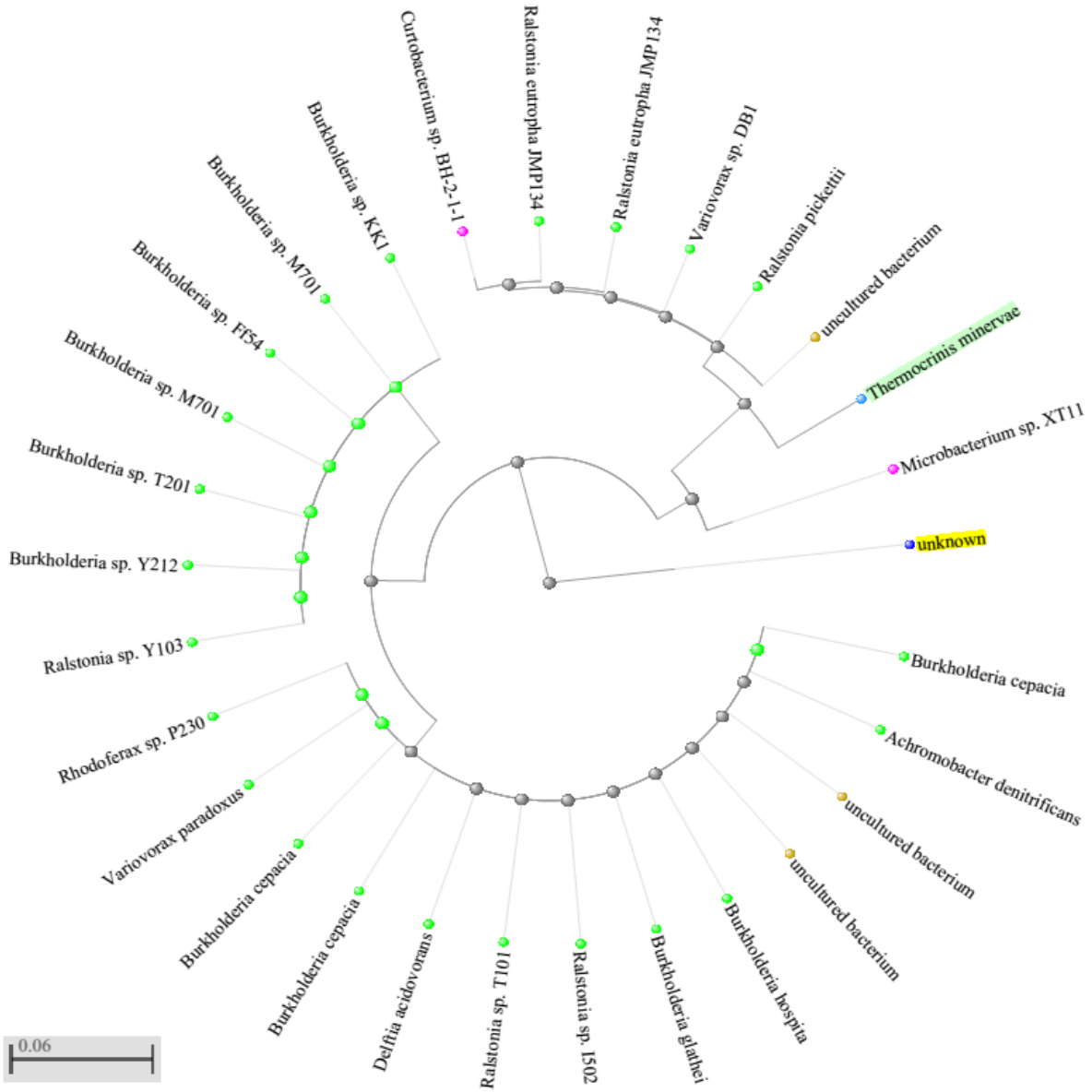


Figure C 6 Neighbour joining phylogenetic tree showing the list of microorganisms that the primer set designed for 2,4-dichlorophenol hydroxylase (tfdB) gene targets. These list are obtained from GeneBank database, using the BLAST (Basic Local Alignment Search Tool) algorithm.

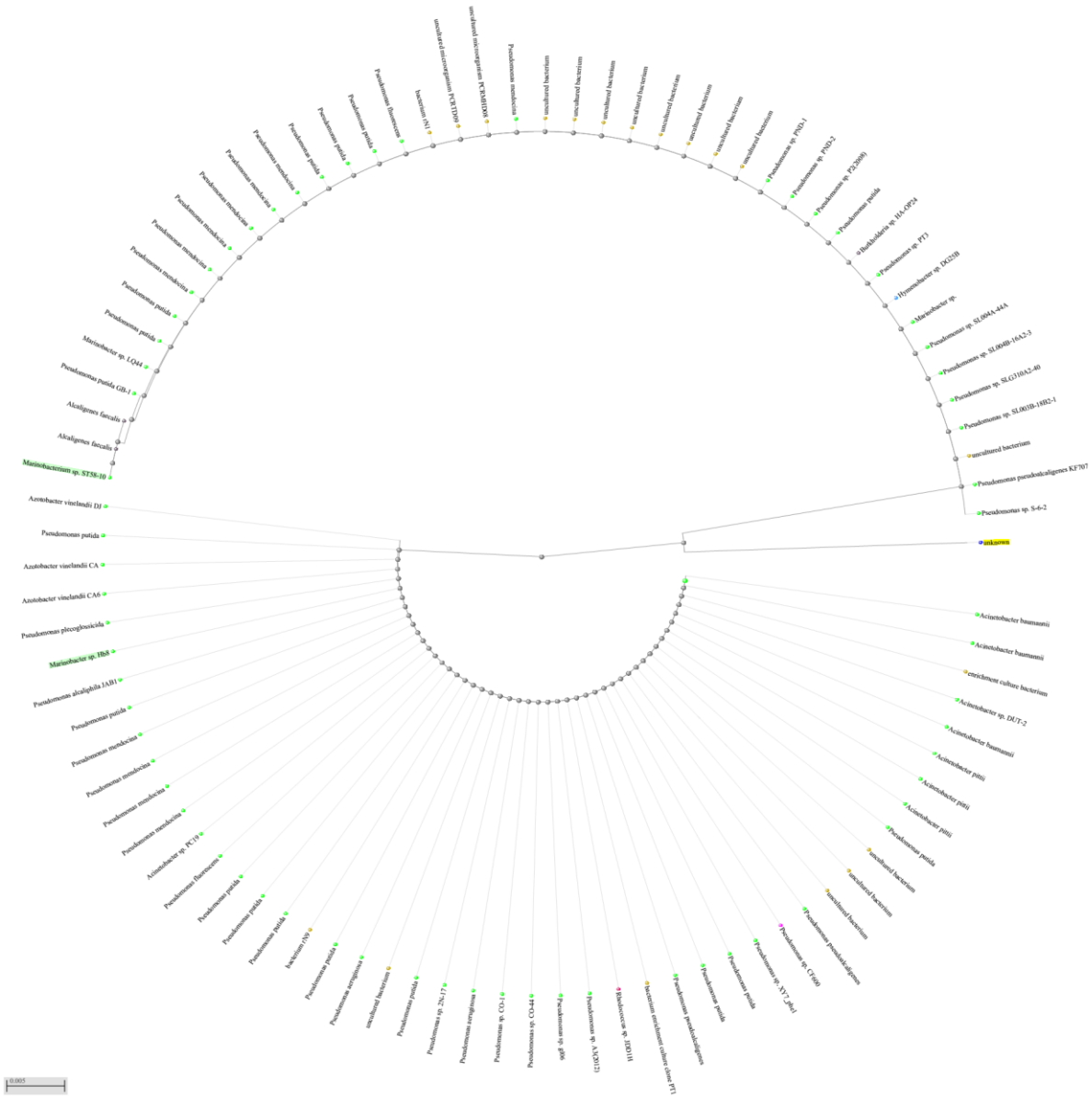


Figure C 7 Neighbour joining phylogenetic tree showing the list of microorganisms that the primer set designed for Phenol hydroxylase (PH) gene targets. These list are obtained from GeneBank database, using the BLAST (Basic Local Alignment Search Tool) algorithm.

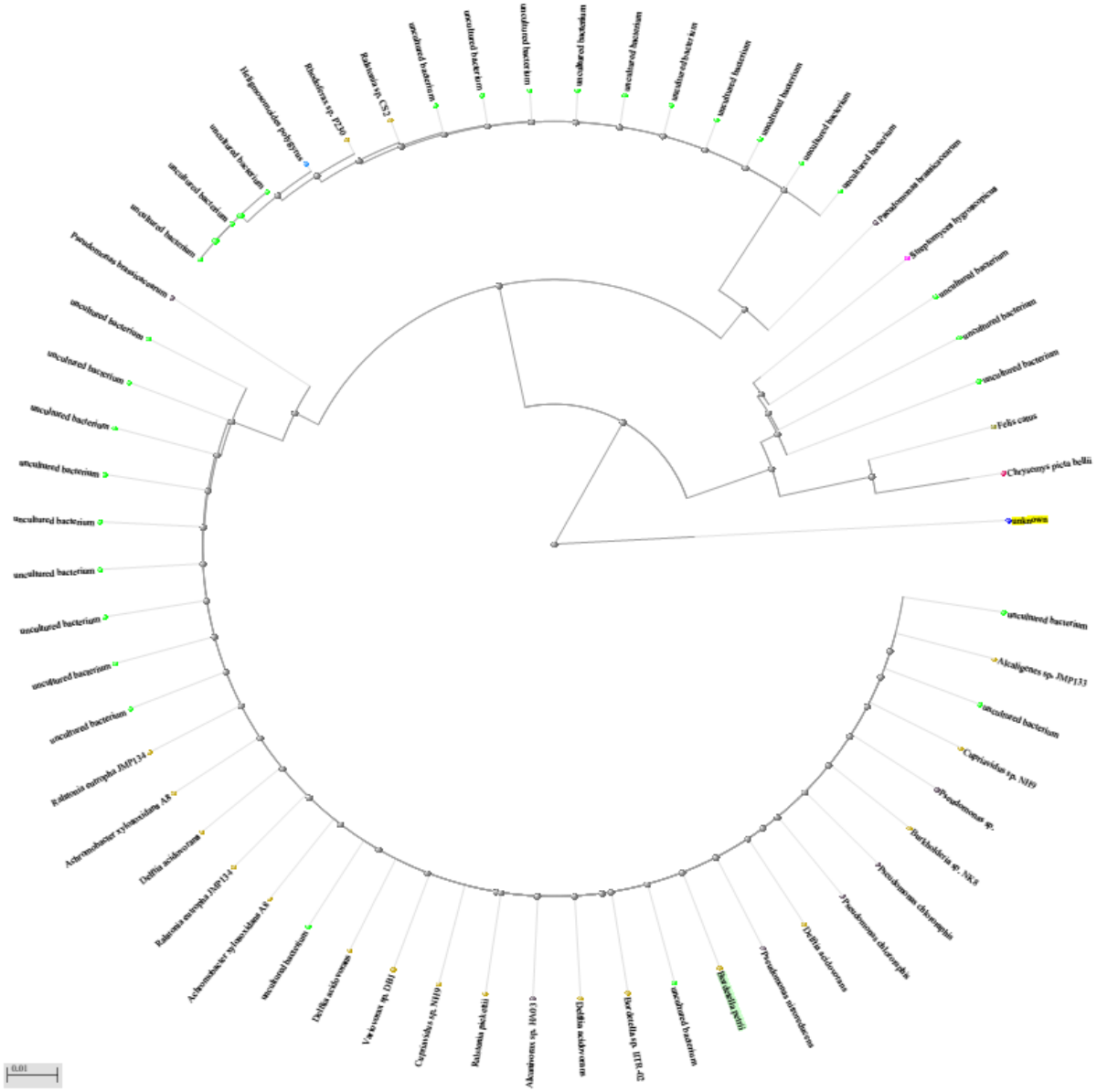


Figure C 8. Neighbour joining phylogenetic tree showing the list of microorganisms that the primer set designed for *tfdC* gene targets. These list are obtained from GeneBank database, using the BLAST (Basic Local Alignment Search Tool) algorithm.

Appendix C.12

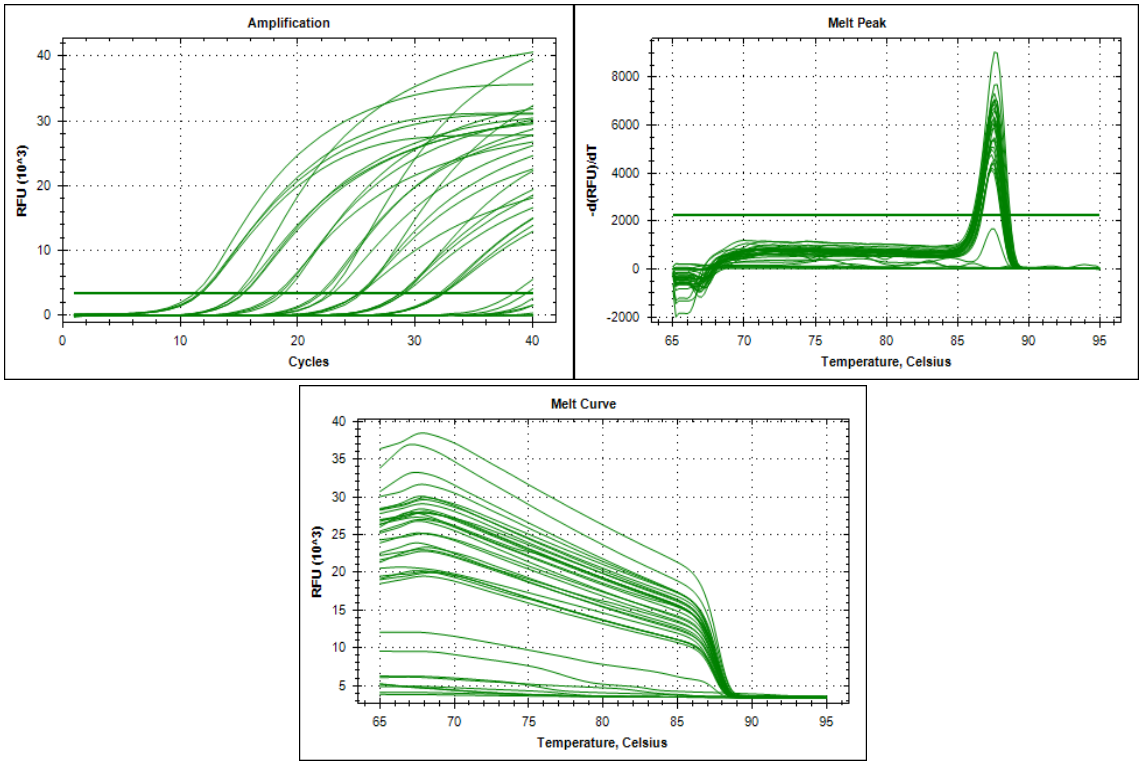


Figure C 9Amplification, melt peak and melt curve charts for catechol-2,3-dioxygenase (C23D) primer set for real time PCR conditions

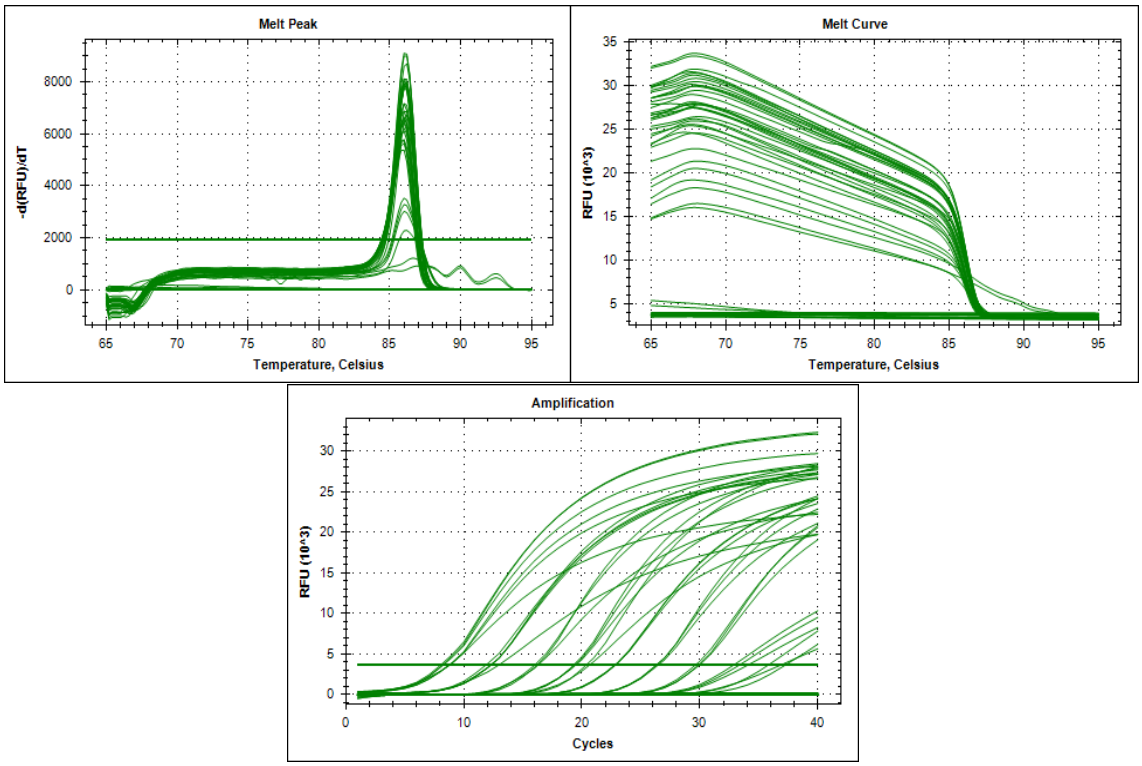


Figure C 10 Melt peak, melt curve and amplification charts for phenol hydroxylase (PH) primer set for real time PCR conditions

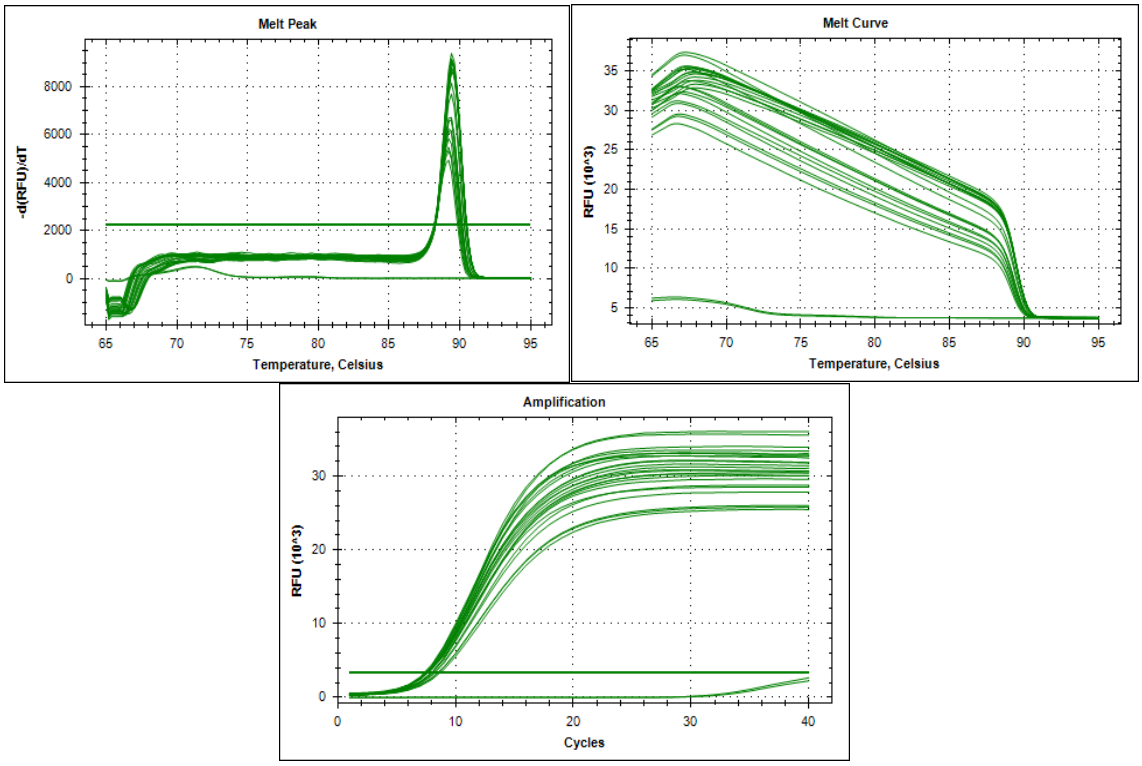


Figure C 11 Melt peak, melt curve and amplification charts for tfdB primer set for real time PCR conditions

Appendix D

1. DOE training experiment for 2,4-DCP

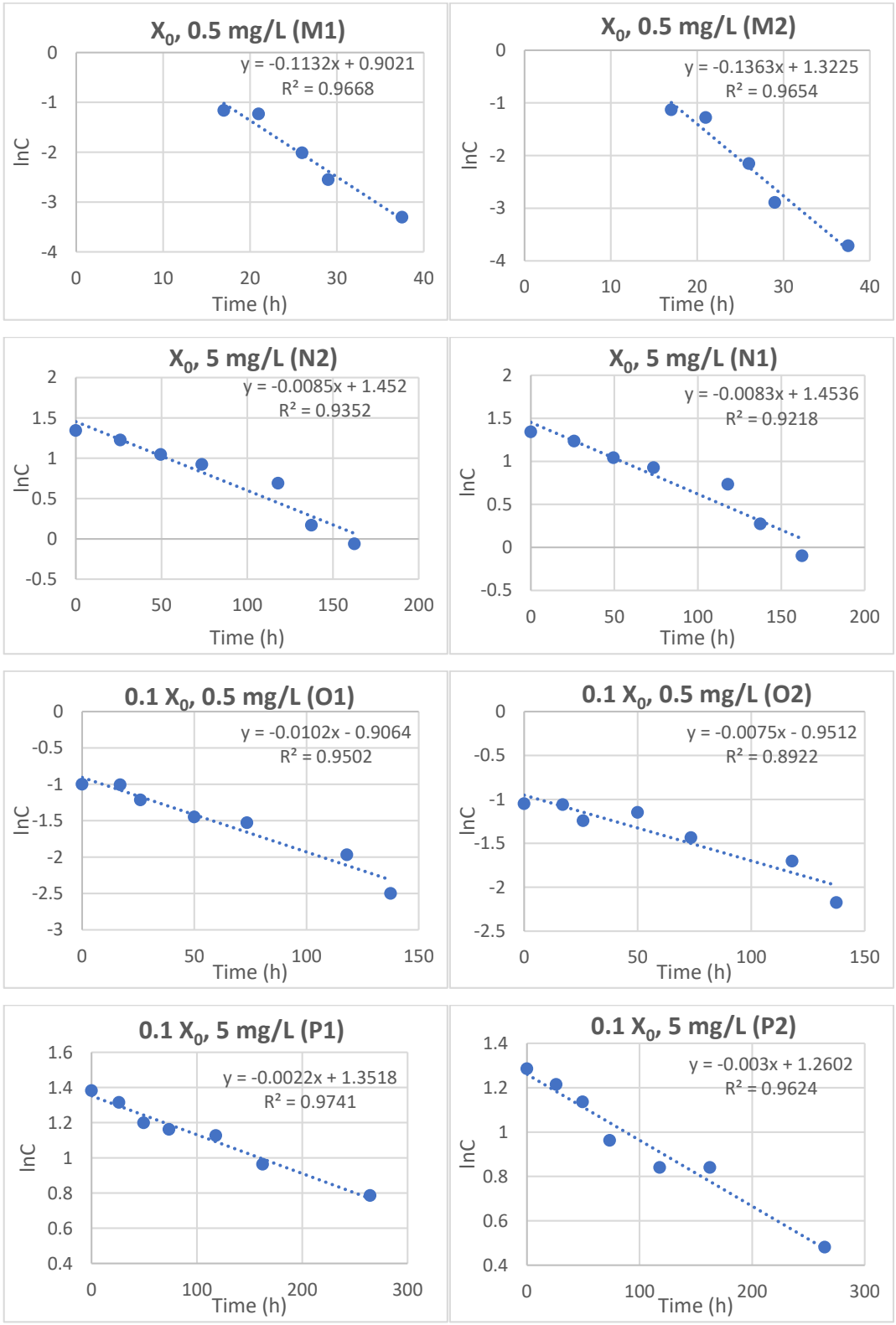


Figure D 1 2,4-DCP biodegradation profile for experimental conditions created with factorial design (2^2) for training experiment set. The figure chart title shows the experimental condition of each bioreactor.

2. DOE 2,4-DCP validation experiment

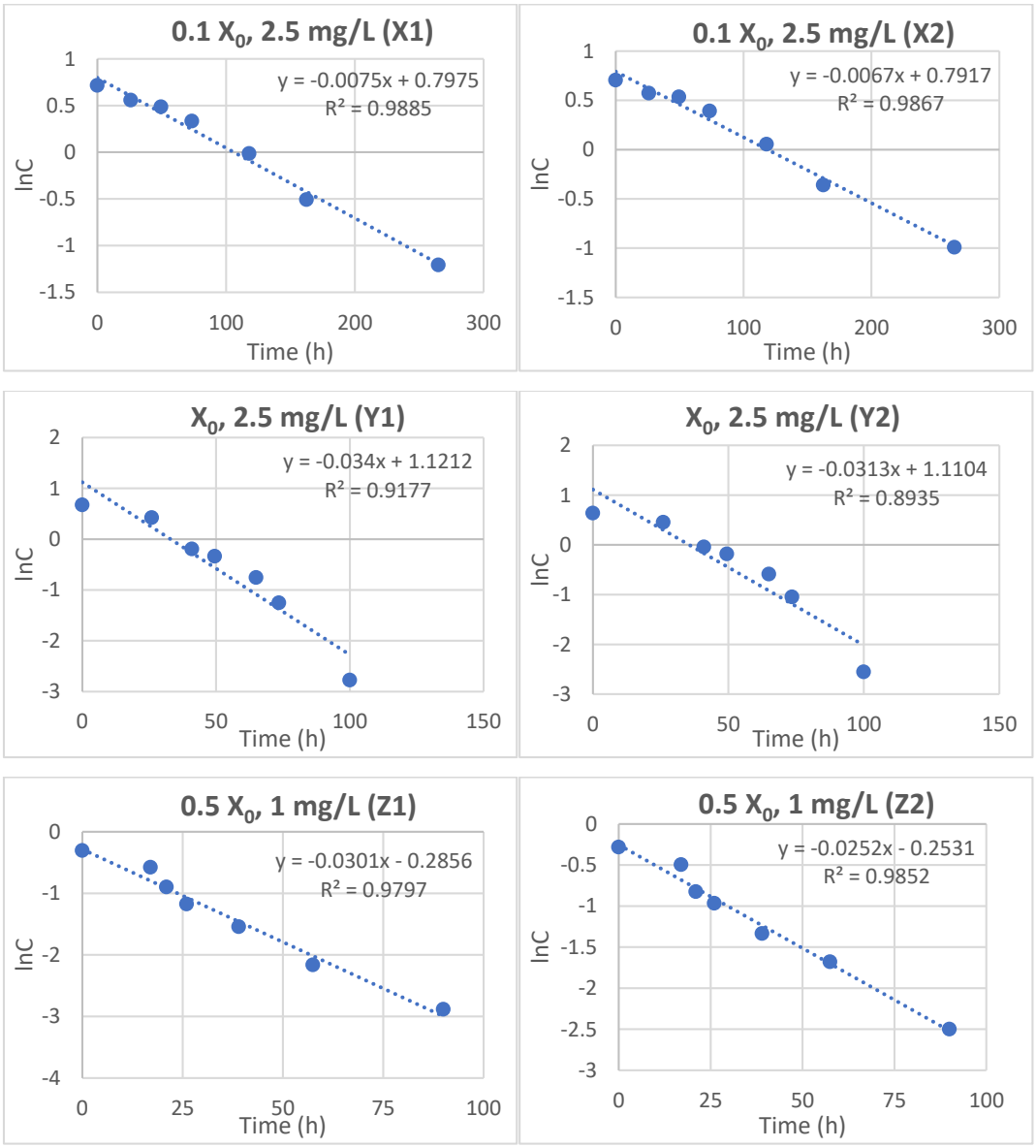


Figure D 2 2,4-DCP biodegradation profiles for DOE validation experiments. The figure chart title shows the experimental condition of each bioreactor.

3. 2,4-DCP Control experiments

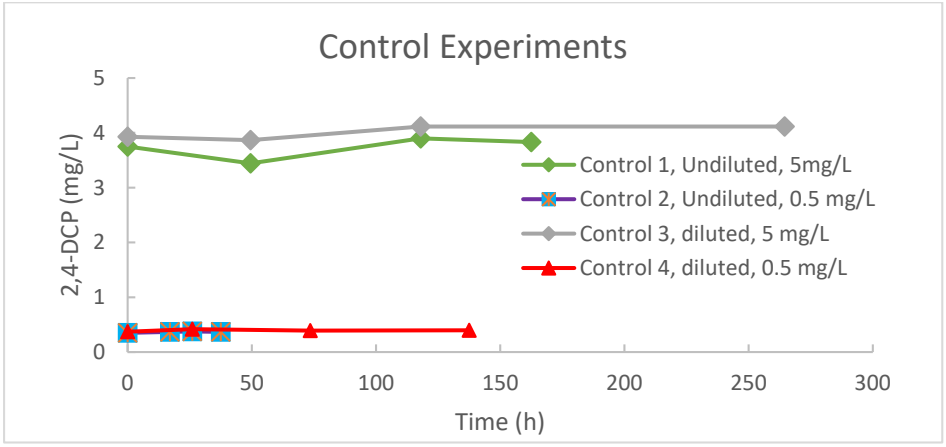


Figure D 3 2,4-DCP biodegradation profiles for control experiments

4. Integrated Monod Equation

$$-\frac{ds}{dt} = u_{max} * S\,(S_0 + X_0 - S)/(K_s + S) \quad (1)$$

Where ds/dt is the rate of substrate removal, u_{max} is the maximum specific growth rate, S is the substrate concentration, S_0 is the initial substrate concentration, X_0 is the initial active degrader abundance and Ks is the half-saturation constant.

Table D 1 Different rates for 2,4-DCP

	First order			**Second order**			**Zero order**		
ID	r2	k	SE	r2	k	SE	r2	k	SE
M1	0.9660	0.1130	0.0121	0.9180	1.2100	0.2090	0.8800	0.0140	0.0031
M2	0.9650	0.1360	0.0140	0.9040	1.9100	0.3500	0.8710	0.0150	0.0035
N1	0.8900	0.0081	0.0012	0.7520	0.0044	0.0011	0.9700	0.0172	0.0013
N2	0.9350	0.0085	0.0010	0.8489	0.0046	0.0009	0.9820	0.0178	0.0010
O1	0.9501	0.0102	0.0010	0.8405	0.0599	0.0116	0.9640	0.0020	0.0002
O2	0.8920	0.0078	0.0011	0.8090	0.0371	0.0080	0.9300	0.0017	0.0002
P1	0.9740	0.0022	0.0002	0.9820	0.0008	0.0000	0.9488	0.0065	0.0007
P2	0.9623	0.0029	0.0003	0.9149	0.0046	0.0006	0.9270	0.0073	0.0009
X1	0.9868	0.0066	0.0003	0.9270	0.0083	0.0010	0.9608	0.0065	0.0006
X2	0.9880	0.0075	0.0036	0.9160	0.0107	0.0015	0.9535	0.0068	0.0007
Y1	0.9177	0.0339	0.0045	0.5980	0.1310	0.0480	0.9370	0.0202	0.0023
Y2	0.8935	0.0312	0.0048	0.5950	0.1040	0.0380	0.9620	0.0195	0.0017
Z1	0.9790	0.0301	0.0019	0.9321	0.1908	0.0230	0.8188	0.0074	0.0016
Z2	0.0251	0.0251	0.0014	0.9169	0.1210	0.0160	0.8570	0.0073	0.0013

Table D 2 Different rates for Phenol

	First order			Second order			Zero order		
I.D	r2	k	SE (k)	R2	k	SE (k)	R2	k	SE (k)
R1	0.966	0.853	0.0715	0.95	2.66	0.26	0.696	0.59	0.175
R2	0.979	1.193	0.0766	0.912	10.36	1.43	0.63	0.57	0.195
B1	0.846	0.065	0.011	0.802	0.0032	0.006	0.877	1.361	0.2
B2	0.942	0.056	0.0057	0.9166	0.0028	0.0003	0.955	1.16	0.102
A1	0.971	3.41	0.29	0.709	10.36	3.31	0.824	2.63	0.609
A2	0.947	3.941	0.535	0.719	8.82	3.18	0.92	3.5	0.58
D1	0.975	0.3364	0.023	0.869	0.032	0.0056	0.943	4.16	0.454
D2	0.947	0.3199	0.033	0.838	0.03	0.006	0.9	3.98	0.58
N	0.879	1.06	0.176	0.66	0.461	0.147	0.99	5.47	0.22
X	0.931	1.188	0.171	0.932	0.79	0.12	0.926	1.76	0.028
Y	0.978	0.785	0.083	0.97	0.36	0.044	0.818	2.97	0.99
Z	0.838	0.665	0.145	0.94	0.279	0.034	0.64	3.51	1.3
M	0.897	1.721	0.291	0.798	1.095	0.27	0.94	3.61	0.44
F	0.967	0.785	0.064	0.88	0.22	0.036	0.95	3.48	0.33
G	0.914	1.14	0.156	0.736	0.299	0.0801	0.977	6.37	0.43
H	0.914	0.606	0.082	0.911	0.095	0.013	0.92	6.21	0.81

Table D 3 Analysis of variance for the multiple regression linear model (Equation 1A) for phenol from the factorial design experiments based of C23D gene copies number.

A)

Source	**DF**	**Adjusted Sum of Squares**	**Adjusted Mean square**	**F-Value**	**p-Value**
Model	3	17.4563	5.81875	1783.88	0.000
Linear	2	14.3771	7.18855	2203.82	0.000
X_0	1	4.4874	4.48741	1375.73	0.000
C_0	1	9.8897	9.88968	3031.92	0.000
2-way interaction	1	3.0792	3.07917	943.99	0.000
X_0*C_0	1	3.0792	3.07917	943.99	0.000
Error	4	0.013	0.00326		
Total	7	17.4693			

B)

Coded Coefficients					
Term	**Effect**	**Coefficient**	**SE Coefficient**	**T-Value**	**P-Value**
Constant		1.2996	0.0202	64.36	0.000
X0	1.4979	0.7490	0.0202	37.09	0.000
C0	-2.2237	-1.1119	0.0202	-55.06	0.000
X0*C0	-1.2408	-0.6204	0.0202	-30.72	0.000

Table D 4 Analysis of variance for multiple regression linear model (equation 2) for 2,4-DCP factorial design experiments based of *tfdC* gene copies number

A)

Source	**DF**	**Adjusted Sum of Squares**	**Adjusted Mean square**	**F-Value**	**p-Value**
Model	3	0.020982	0.006994	103.31	0
Linear	2	0.014921	0.00746	110.2	0
X0	1	0.007405	0.007405	109.39	0
C0	1	0.007515	0.007515	111.01	0
2-way interaction	1	0.006061	0.006061	89.53	0.0001
X0*C0	1	0.006061	0.006061	89.53	0.0001
Error	4	0.000271	0.000068		
Total	7	0.0021253			

B)

Coded Coefficients					
Term	**Effect**	**Coefficient**	**SE Coefficient**	**T-Value**	**P-Value**
Constant		0.03615	0.00291	12.43	0
X0	0.06085	0.03043	0.00291	10.46	0
C0	-0.06130	-0.03065	0.00291	-10.54	0
X0*C0	-0.05505	-0.02752	0.00291	-9.46	0.001

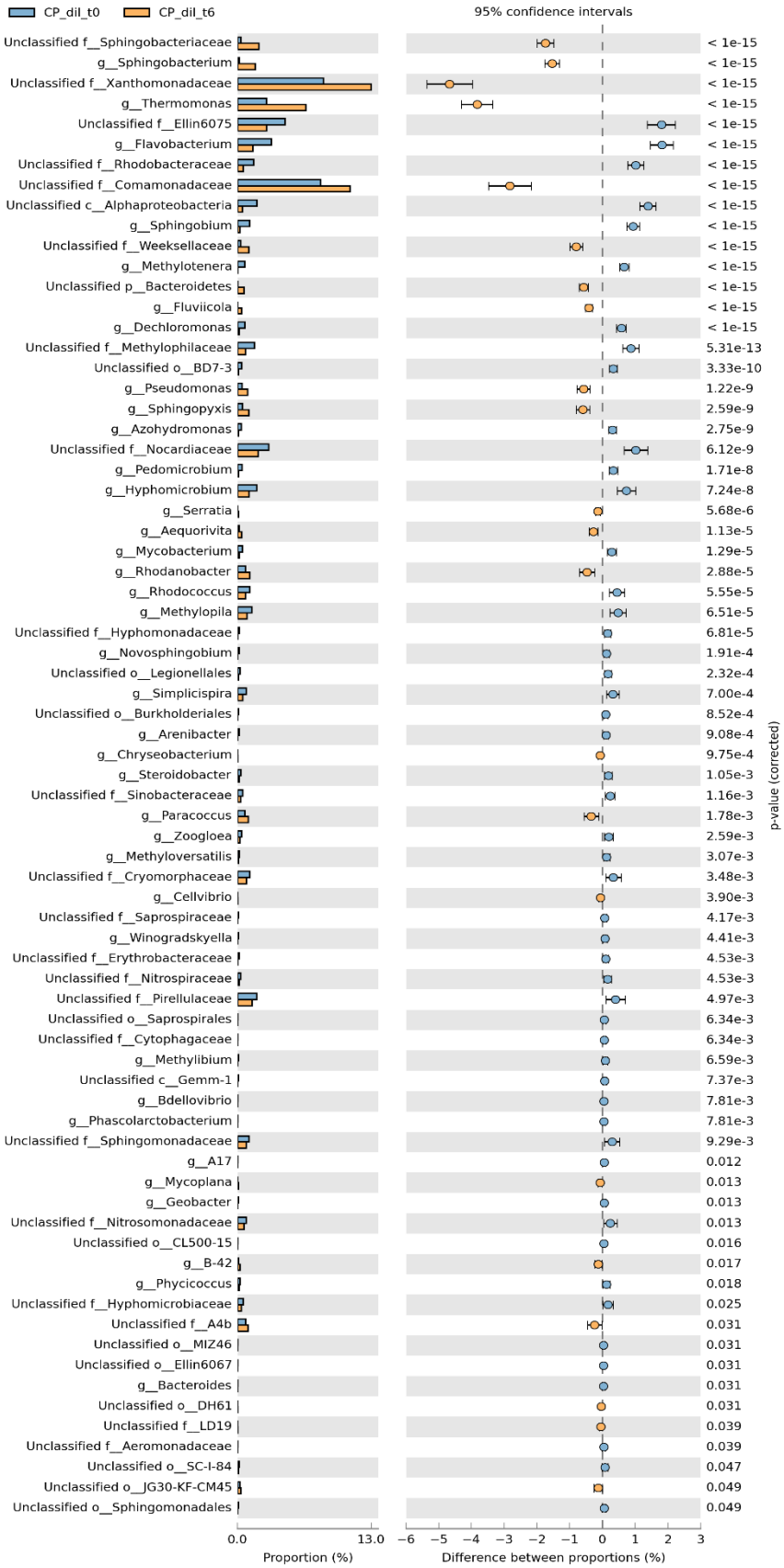


Figure D 4 Extended error bar plot considering abundance profile of microbial taxa in 16S rRNA amplicon sequencing data for 4-Chlorophenol degradation assay in diluted inoculum. The bacterial domain that increased or decreased significantly (G-test, p<0.05) over the duration of assay are reported in the plot.

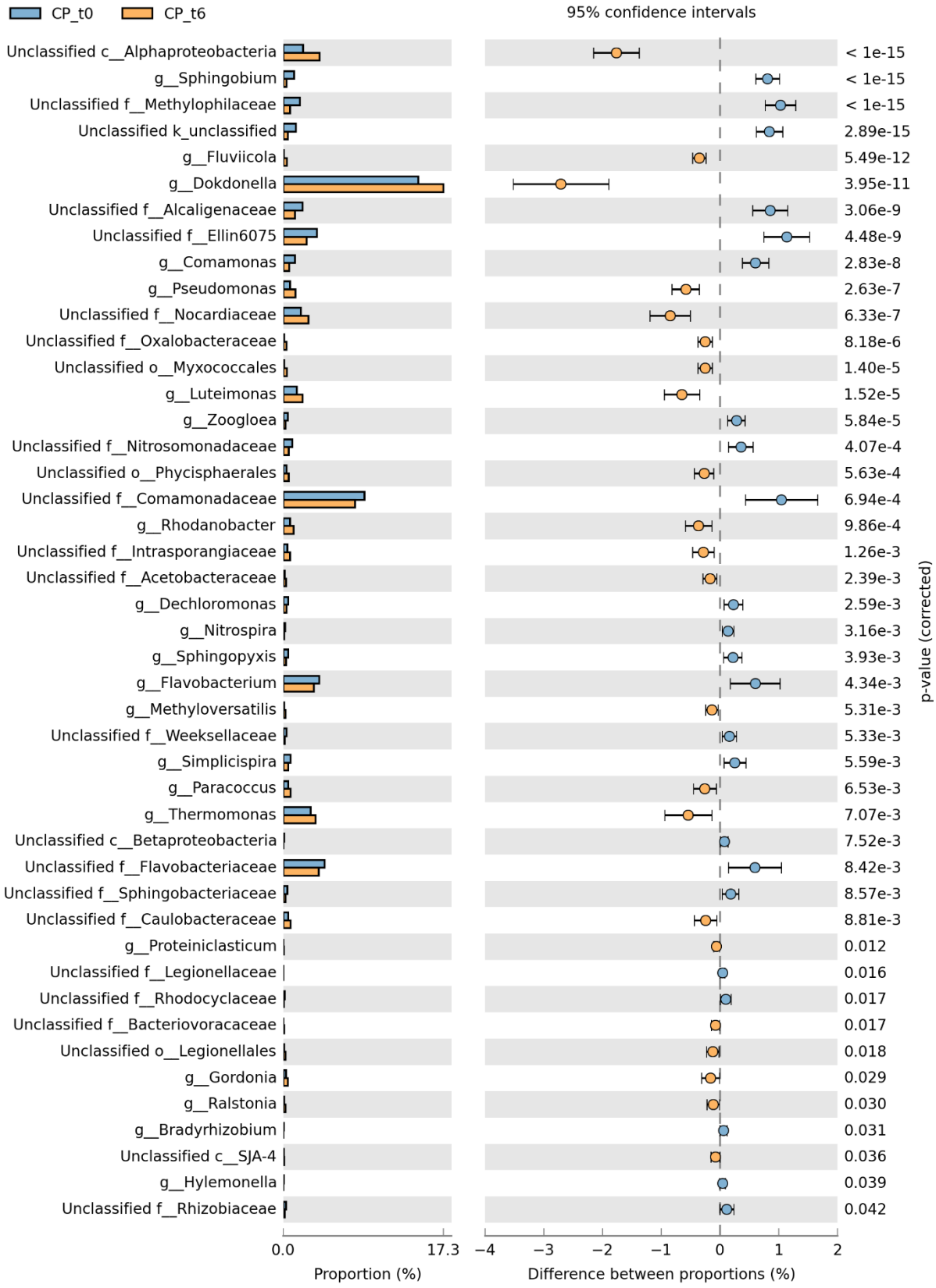


Figure D 5 Extended error bar plot considering abundance profile of microbial taxa in 16S rRNA amplicon sequencing data for 4-Chlorophenol degradation assay in undiluted inoculum. The bacterial domain that increased or decreased significantly (G-test, p<0.05) over the duration of assay are reported in the plot

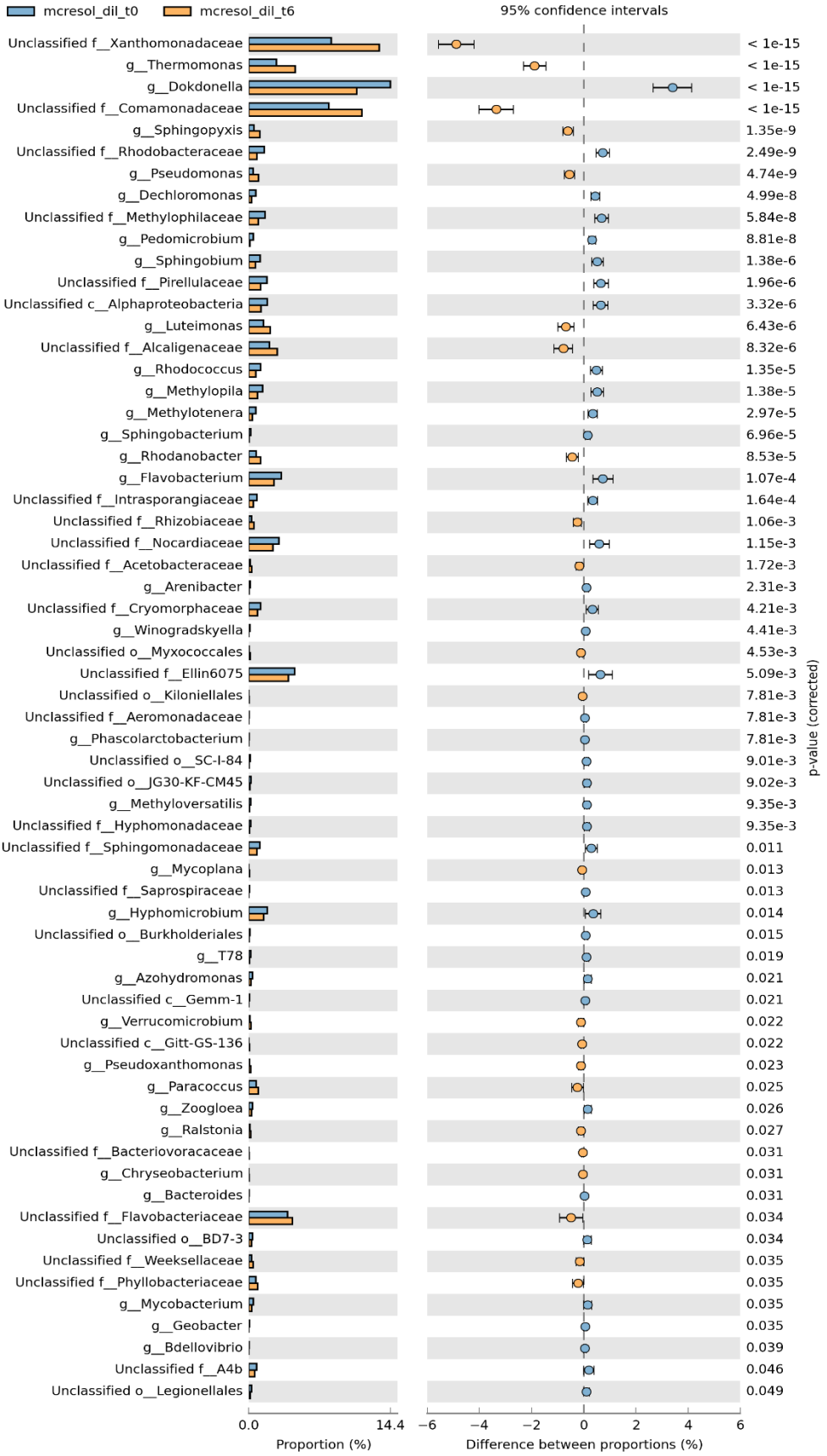


Figure D 6 Extended error bar plot considering abundance profile of microbial taxa in 16S rRNA amplicon sequencing data for m-cresol degradation assay in diluted inoculum. The bacterial domain that increased or decreased significantly (G-test, p<0.05) over the duration of assay are reported in the plot

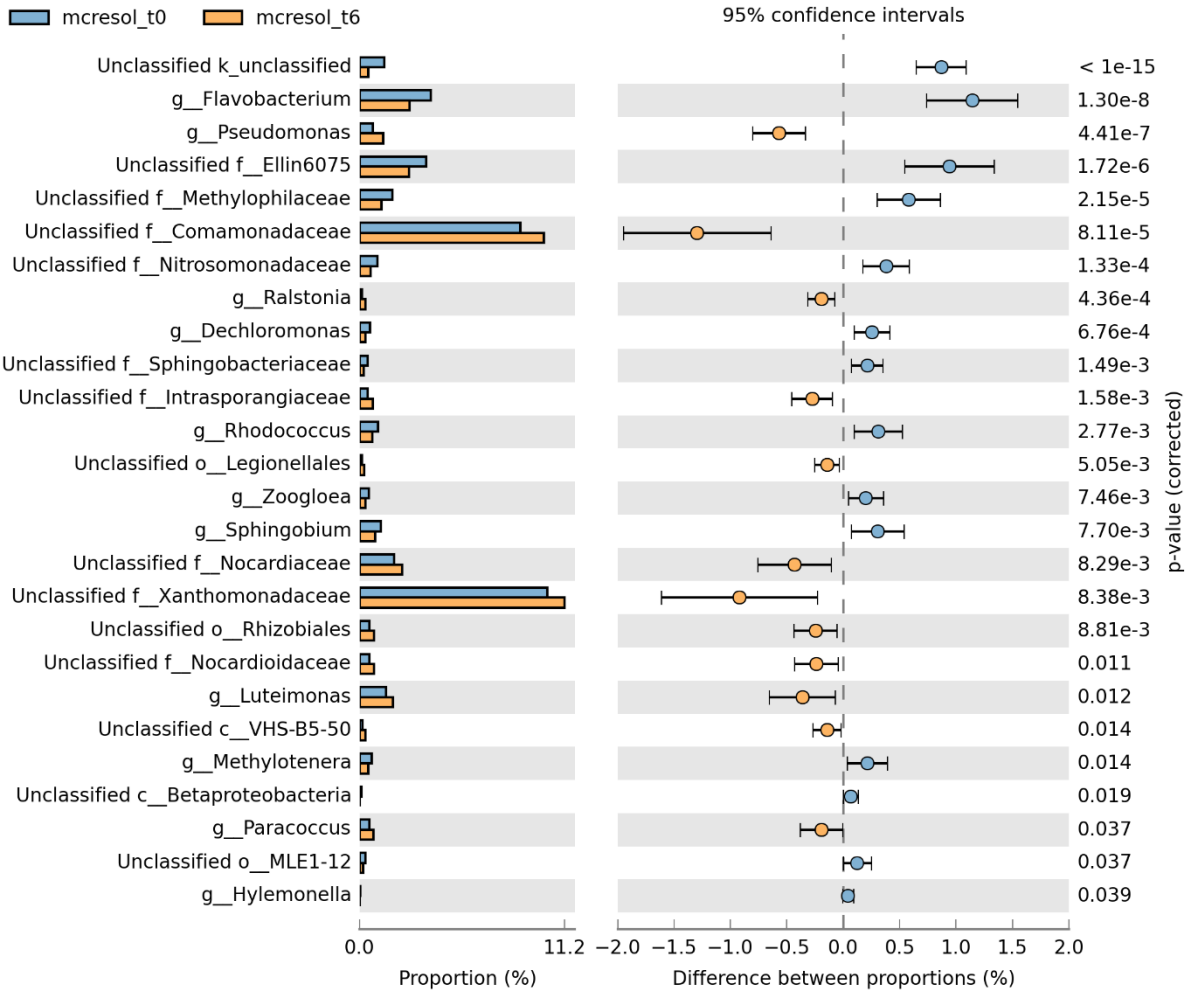


Figure D 7 Extended error bar plot considering abundance profile of microbial taxa in 16S rRNA amplicon sequencing data for m-cresol degradation assay in undiluted inoculum. The bacterial domain that increased or decreased significantly (G-test, p<0.05) over the duration of assay are reported in the plot

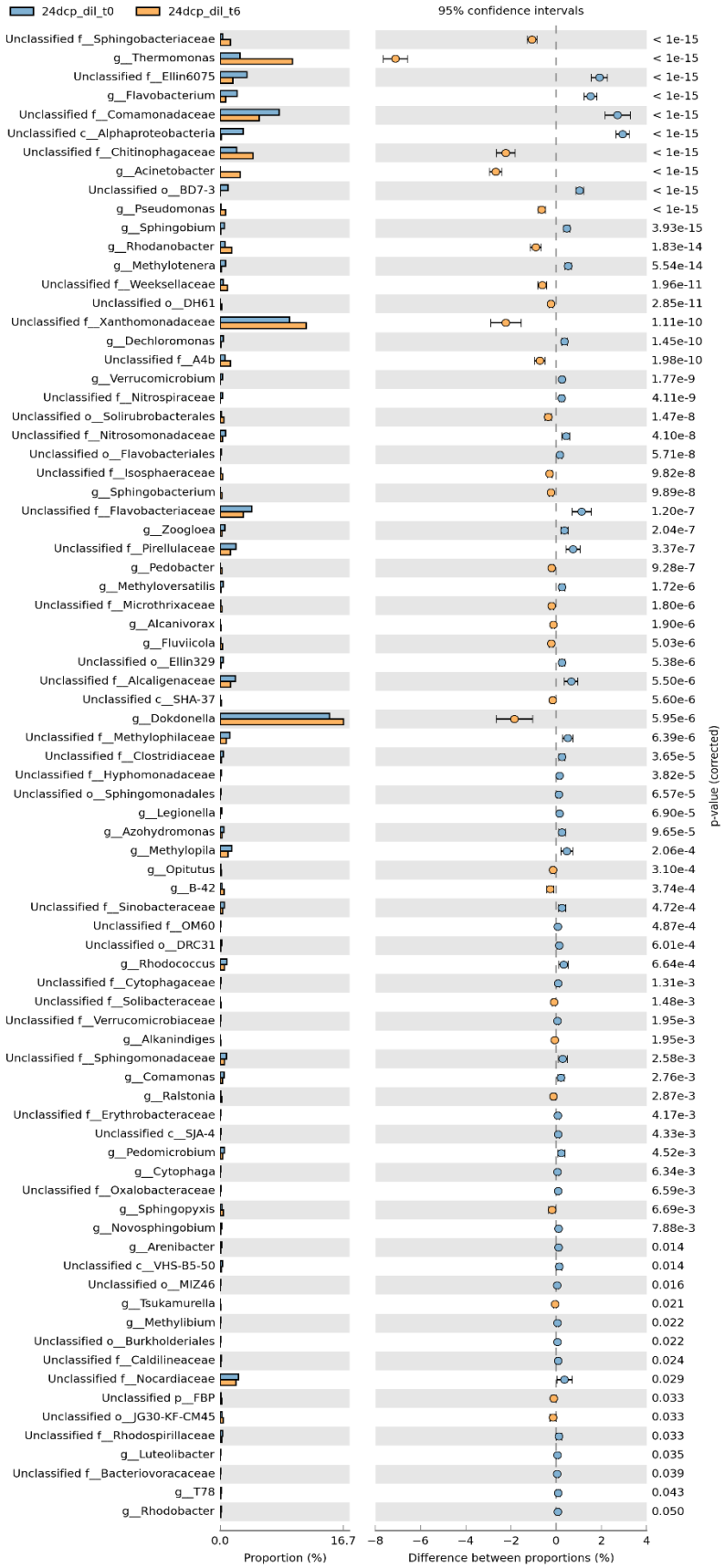


Figure D 8 Extended error bar plot considering abundance profile of microbial taxa in 16S rRNA amplicon sequencing data for 2,4-DCP degradation assay in diluted inoculum. The bacterial domain that increased or decreased significantly (G-test, $p<0.05$) over the duration of assay are reported in the plot

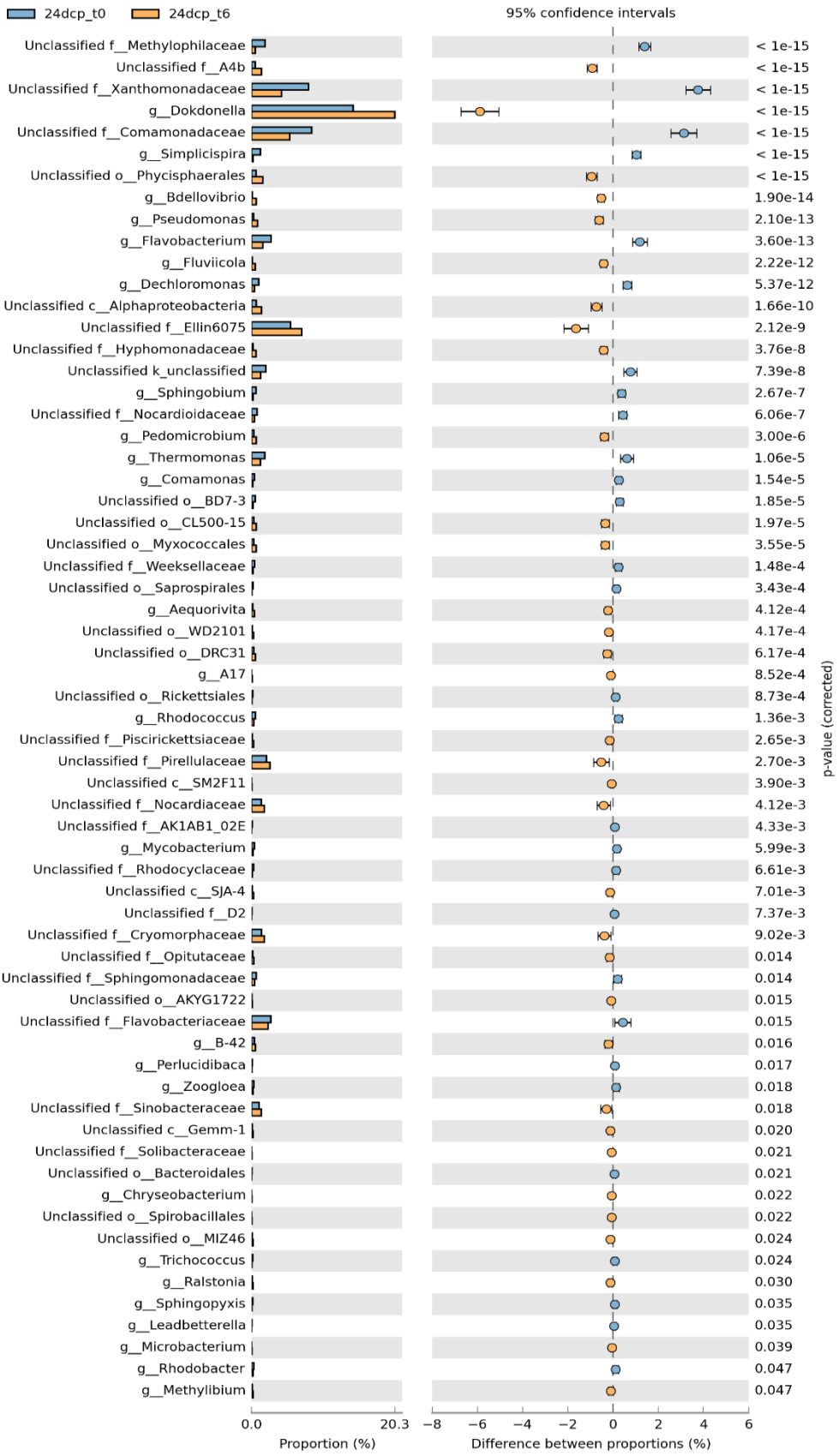


Figure 23 Extended error bar plot considering abundance profile of microbial taxa in 16S rRNA amplicon sequencing data for 2,4-DCP degradation assay in undiluted inoculum. The bacterial domain that increased or decreased significantly (G-test, p<0.05) over the duration of assay are reported in the plot